bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1 **Functional genomics of the stable fly, *Stomoxys calcitrans*, reveals mechanisms**
2 **underlying reproduction, host interactions, and novel targets for pest control**
3

4 Pia U. Olafson[1,*], Serap Aksoy[2], Geoffrey M. Attardo[3], Greta Buckmeier[1], Xiaoting Chen[4], Craig
5 J. Coates[5], Megan Davis[1], Justin Dykema[6], Scott J. Emrich[7], Markus Friedrich[6], Christopher J.
6 Holmes[8], Panagiotis Ioannidis[9], Evan N. Jansen[8], Emily C. Jennings[8], Daniel Lawson[10], Ellen O.
7 Martinson[11], Gareth L. Maslen[10], Richard P. Meisel[12], Terence D. Murphy[13], Dana Nayduch[14],
8 David R. Nelson[15], Kennan J. Oyen[8], Tyler J. Raszick[5], José M. C. Ribeiro[16], Hugh M.
9 Robertson[17], Andrew J. Rosendale[18], Timothy B. Sackton[19], Sonja L. Swiger[20], Sing-Hoi Sze[21],
10 Aaron M. Tarone[5], David B. Taylor[22], Wesley C. Warren[23], Robert M. Waterhouse[24], Matthew T.
11 Weirauch[25] John H. Werren[26], Richard K. Wilson[27], Evgeny M. Zdobnov[9], and Joshua B.
12 Benoit[8,*]
13

14 [1]Livestock Arthropod Pests Research Unit, USDA-ARS, Kerrville TX, USA
15 [2]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven
16 CT, USA
17 [3]Department of Entomology and Nematology, University of California - Davis, Davis CA, USA
18 [4]The Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical
19 Center, Cincinnati, OH, USA
20 [5]Department of Entomology, Texas A & M University, College Station, TX, USA
21 [6]Department of Biological Sciences, Wayne State University, Detroit, MI, USA
22 [7]Department of Electrical Engineering & Computer Science, University of Tennessee, Knoxville
23 TN, USA
24 [8]Department of Biological Sciences, University of Cincinnati, Cincinnati, OH, USA
25 [9]Department of Genetic Medicine and Development, University of Geneva Medical School and
26 Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland
27 [10] The European Molecular Biology Laboratory, The European Bioinformatics Institute, The
28 Wellcome Genome Campus, Hinxton, CB10 1SD, United Kingdom
29 [11]Department of Entomology, University of Georgia, Athens, GA, USA
30 [12]Department of Biology and Biochemistry, University of Houston, Houston, TX, USA
31 [13]National Center for Biotechnology Information, National Library of Medicine, National Institutes
32 of Health, Bethesda, Maryland, USA
33 [14]Arthropod-borne Animal Diseases Research Unit, USDA-ARS, Manhattan, KS, USA
34 [15]University of Tennessee Health Science Center, Memphis, TN, USA

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

35 [16]Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of
36 Allergy and Infectious Diseases, Rockville, MD, USA
37 [17]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL, USA
38 [18] Department of Biology, Mount St. Joseph University, Cincinnati, OH, USA
39 [19]Informatics Group, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, USA
40 [20]Department of Entomology, Texas A&M AgriLife Research and Extension Center,
41 Stephenville, TX, USA
42 [21]Department of Computer Science & Engineering, Department of Biochemistry & Biophysics,
43 Texas A & M University, College Station, TX, USA
44 [22]Agroecosystem Management Research Unit, USDA-ARS, Lincoln, NE, USA
45 [23]University of Missouri, Bond Life Sciences Center, Columbia, MO, USA
46 [24]Department of Ecology and Evolution, University of Lausanne, and Swiss Institute of
47 Bioinformatics, 1015 Lausanne, Switzerland
48 [25]The Center for Autoimmune Genomics and Etiology, Division of Biomedical Informatics and
49 Developmental Biology, Cincinnati Children's Hospital Medical Center, Department of
50 Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA
51 [26]Department of Biology, University of Rochester, Rochester, NY, USA
52 [27]Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA; Ohio
53 State University, College of Medicine, Columbus OH, USA
54 *Corresponding authors
55

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

56 **Abstract**

57

58 **Background:** The stable fly, *Stomoxys calcitrans*, is a major blood-feeding pest of livestock that
59 has near worldwide distribution, causing an annual cost of over $2 billion for control and product
60 loss in the United States alone. Control of these flies has been limited to increased sanitary
61 management practices and insecticide application for suppressing larval stages. Few genetic
62 and molecular resources are available to help in developing novel methods for controlling stable
63 flies.

64

65 **Results:** This study examines stable fly biology by utilizing a combination of high-quality
66 genome sequencing, microbiome analyses, and RNA-seq analyses targeting multiple
67 developmental stages and tissues. In conjunction, manual curation of over 1600 genes was
68 used to examine gene content related to stable fly reproduction, interactions with their host,
69 host-microbe dynamics, and putative routes for control. Most notable was establishment of
70 reproduction-associated genes and identification of expanded vision, chemosensation, immune
71 repertoire, and metabolic detoxification pathway gene families.

72

73 **Conclusions:** The combined sequencing, assembly, and curation of the male stable fly genome
74 followed by RNA-seq and downstream analyses provide insights necessary to understand the
75 biology of this important pest. These resources and knowledge will provide the groundwork for
76 expanding the tools available to control stable fly infestations. The close relationship of
77 *Stomoxys* to other blood-feeding (*Glossina*) and non-blood-feeding flies (medflies, *Drosophila*,
78 house flies) will allow for understanding the evolution of blood feeding among Cyclorrhapha
79 flies.

80

81 **Keywords**: stable fly genome, muscid genomics, insect orthology, chemoreceptor genes, opsin
82 gene duplication, insecticide resistance genes, insect adaptation, gene regulation, insect
83 immunity

84 **Additional File 1:** Supplementary Material; **Additional File 2:** Supplementary Tables

85 **Supplementary Dataset Files**: 10

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

86 **INTRODUCTION**

87 Livestock ectoparasites are detrimental to cattle industries in the US and worldwide, impacting
88 both confined and rangeland operations. Flies from the Muscidae family commonly occupy
89 these settings, including the nonbiting house fly and face fly and the blood-feeding
90 (hematophagous) stable fly and horn fly. These muscid flies exhibit different larval and adult
91 biologies, varying in larval developmental substrates, as well as adult nutrient sources and
92 feeding frequency [1, 2]. As such, control efforts against these flies are not one size fits all. The
93 stable fly, *Stomoxys calcitrans* (L.), in particular, is a serious hematophagous pest with a
94 cosmopolitan host range, feeding on bovids, equids, cervids, canines, and occasionally humans
95 throughout much of the world. The stable fly's painful bites disrupt livestock feeding behavior [3-
96 6]; these bites can be numerous during heavy infestation, leading to reductions of productivity
97 by over $2 billion USD [7]. In Australia, Brazil, and Costa Rica, dramatic increases in stable fly
98 populations have coincided with the expansion of agricultural production where the vast
99 accumulation of post-harvest byproducts are recognized as nutrient sources for development of
100 immature stages [8-10].

101 Stable fly larvae occupy and develop in almost any type of decomposing vegetative
102 materials, e.g. spent hay, grass clippings, residues from commercial plant processing, that are
103 often contaminated with animal wastes [11]. The active microbial communities residing in these
104 developmental substrates, e.g. plant, soil, manure, are required for larval development and
105 likely provide essential nutrients [12]. Even though stable flies are consistently exposed to
106 microbes during feeding and grooming activities, biological transmission (uptake, development,
107 and subsequent transmission of a microbial agent by a vector) of pathogens has not been
108 demonstrated for organisms other than the helminth *Habronema microstoma* [13]. Stable flies
109 have been implicated in mechanical transmission (transfer of pathogens from an infected host
110 or a contaminated substrate to a susceptible host, association between specific vector and
111 pathogen is not necessary) of Equine infectious anemia, African swine fever, West Nile, and Rift
112 Valley Viruses, *Trypanosoma* spp., and *Besnoitia* spp. (reviewed by [13]). The apparent low
113 vector competence of stable flies implicates the importance of immune system pathways not
114 only in regulating larval survival in microbe-rich environments but also in the inability of
115 pathogens to survive and replicate in the adult midgut following ingestion [14-16].

116 Stable fly mate location and recognition are largely dependent upon visual cues and contact
117 pheromones [17, 18], and gravid females identify suitable oviposition sites through a
118 combination of olfactory and contact chemostimuli along with physical cues [19, 20]. Since
119 stable flies infrequently associate with their hosts, feeding only 1 to 2 times per day, on-animal

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

120 and pesticide applications are less effective control efforts than those that integrate sanitation
121 practices with fly population suppression by way of traps [21]. Given the importance of
122 chemosensory and vision pathways, repellents have been identified that target stable fly
123 chemosensory inputs and current trap technologies exploit stable fly visual attraction [22-24].
124 However, despite these efforts, consistent control of stable fly populations remains challenging
125 and development of novel control mechanisms is greatly needed.
126 Although both sexes feed on sugar, adults are reliant on a bloodmeal for yolk deposition and
127 egg development, as well as seminal fluid production [25, 26]. Blood feeding evolved
128 independently on at least five occasions within the Diptera, in the Culicimorpha,
129 Psychodomorpha, Tabanomorpha, Muscoidea, and Hippoboscoidea [27]. The Muscinae appear
130 to have a high propensity for developing blood feeding; which has occurred at least four times
131 within this subfamily - once in each of the *domestica*-, *sorbens*- and *lusoria*- groups and again in
132 the Stomoxini [28]. Unlike other groups of blood-feeding Diptera where non-blood feeding
133 ancestors are distantly related and / or difficult to discern, stomoxynes are imbedded with the
134 subfamily Muscinae of the Muscidae, featuring many non-blood feeding species. Contrasting
135 blood-feeding culicimorphs and tabanimorphs, stable flies exhibit gonotrophic discordance [29,
136 30], requiring 3-4 blood meals for females to develop their first clutch of eggs and an additional
137 2-3 for each subsequent clutch of eggs. These unique aspects of stable flies offer opportunities
138 for comparative analysis of the genomic features underlying these key biological traits.
139 Even with the importance of the stable fly as a pest, little is known about the molecular
140 mechanisms underlying the biology of *S. calcitrans*. To further our understanding of this critical
141 livestock pest, we report a draft genome sequence of the stable fly. The quality of this genome
142 is high and includes *in silico* annotation that was aided by extensive developmental and tissue-
143 specific RNA-seq data focusing on the feeding and reproduction of *S. calcitrans*. Manual
144 curation and comparative analyses focused on aspects related to host interactions,
145 reproduction, control, and regulation of specific biological processes. Our study significantly
146 advances the understanding of stable fly biology including the identification of unique molecular
147 and physiological processes associated with this blood-feeding fly. These processes can serve
148 as novel targets which will assist in both developing and improving control of this important
149 livestock pest.
150

151 ## RESULTS AND DISCUSSION

152

153 ### Genome assembly and annotation supported by comparative and functional genomics

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

154 Whole genome shotgun sequencing of adult males resulted in the 66x coverage draft
155 assembly of 971 MB of total sequence. Scaffolds (12,042) and contigs (125,702) had N50
156 lengths of 504.7 and 11.3 kb, respectively. The sequence was ~20% smaller compared to the
157 genome predicted by propidium iodide analyses (~1150 MB, [31]). This difference is likely the
158 result of heterochromatin and other repetitive regions that were unassembled, as genome size
159 is not significantly different between the sexes [31], and is comparable to differences
160 documented for other insect genomes [32-34]. Further details of the stable fly genome assembly
161 and analyses are provided as supplementary information (Additional Files 1 and 2). There were
162 16,102 predicted genes/pseudogenes that included 2,003 non-protein coding genes, and a total
163 of 22,450 mRNA transcripts were predicted with over 90-95% supported by RNA-seq (Additional
164 File 2:Table S2). Manual curation and analyses allowed preliminary chromosome arm
165 assignment and identification of repeat elements from the genome (Additional File 1, Sections 4
166 and 5) and included the combined analyses and correction of over 1,600 genes focused on
167 gene families underlying reproduction, immunity, host sensing, feeding, and insecticide
168 resistance.

169 Completeness of the genome was assessed through identification of sets of benchmarking
170 universal single-copy orthologs (BUSCOs) among flies, and BUSCOs identified were
171 comparable to those in other flies for both the genome and predicted gene set (Additional File 1:
172 Fig. S1). Further, the number of genes with significant alignment to *Drosophila* spp. genes was
173 comparable to other published fly genomes (Fig. 1). Lastly, CEGMA genes and those
174 associated with autophagy were all identified from the *S. calcitrans* genome (Additional File 2,
175 Tables S3 and S4), which can be used as an additional metric of genome completeness as
176 these are highly conserved among flies [32, 35]. These metrics indicate that the genome is of
177 sufficient quality for subsequent comparative analyses with other insects. Comparison of protein
178 orthologs revealed only 47 *Stomoxys* species-specific protein families relative to other higher
179 flies (Fig. 1). Based on gene ontology, there was enrichment for zinc finger transcription factors
180 in relation to all genes, which has also been documented in other insect systems [36, 37].

181 RNA sequencing produced a comprehensive catalog of expression profiles for all genes
182 sampling both sexes, as well as different developmental stages, tissues, and specific organs
183 (Additional File 2: Tables S5 – S13). Specifically, RNA collected from whole females (teneral
184 and mated, reproductive), whole males (teneral), male reproductive tracts (mated), female
185 reproductive tracts (mated), male heads (fed, mated), female heads (fed, mated), third instar
186 larva, and pooled female/male salivary glands were examined to assist in addressing core
187 questions of this study. The RNA-seq datasets were validated by quantitative RT-PCR on a

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

188 sample of 25 genes with a Pearson's correlation of 0.85 (Additional File 1: Section 3; Additional
189 File 2: Table S14). RNA-seq datasets are discussed below in relation to the targeted areas of
190 focus.
191

192 **Unique duplications in the yolk protein gene family and evidence of male-biased genes**
193 **with putative seminal fluid function**

194 Yolk proteins (YPs) in flies act as a primary source of nutrients for developing embryos.
195 These proteins function as a source of amino acids and also likely function as transporters of
196 other essential nutrients such as lipids and vitamins [38]. While many insects utilize YPs
197 classified as vitellogenins, cyclorrhaphan (higher) flies such as *Stomoxys* [39], *Drosophila* [40],
198 *Musca* [41], *Calliphora* [42], and *Glossina* [43] utilize an alternative class of proteins derived
199 from lipase enzymes termed YPs [44]. The number of yolk protein genes varies among species.
200 The species-specific expansions/contractions observed within this class of genes may reflect
201 reproductive demand within those species. Our analysis identified 8 putative *Stomoxys* YP
202 homologs relative to YP gene family members in *Musca* [36], four of which we identified during
203 our analysis.

204 To understand the evolutionary relationships between characterized YPs we performed a
205 phylogenetic analysis of the predicted YPs from *S. calcitrans*, *M. domestica*, *G. morsitans* and
206 *Drosophila melanogaster* (Fig. 2). Based on this analysis, the yolk protein gene family expanded
207 in *S. calcitrans* and *M. domestica* sometime after their divergence from *Drosophila*. Of the
208 *Musca* and *Stomoxys* specific YPs, 3 members are orthologous between the two species
209 suggesting derivation from a common ancestor. However, the remaining yolk protein genes are
210 paralogous and appear to originate from independent duplication events occurring after the
211 divergence of the *Stomoxys* and *Musca* lineages. The lineage-specific expansions suggest that
212 duplications in this gene family may confer a reproductive advantage by increasing reproductive
213 capacity. In support of this role, all 8 YP genes are upregulated in reproductively active *S.*
214 *calcitrans* females. Expression is not observed in the female reproductive tract, which suggests
215 these genes are expressed and translated in the fat body, secreted into the hemolymph and
216 transported to the ovaries as observed in other higher flies (Fig. 2).

217 To identify male-specific reproductive genes and putative seminal proteins we performed an
218 analysis comparing RNA-seq data derived from male and female reproductive tract tissues (Fig.
219 3). Genes with a male reproductive expression of at least 50 reads per kilobase of transcript per
220 million mapped reads (RPKM) or higher and that were expressed at least 5-fold higher in males

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

221 relative to females were selected as male biased. This analysis resulted in the classification of
222 763 genes with male reproductive tract biased expression (Additional File 2: Table S18).
223 We performed a reciprocal BLAST analysis to identify orthologs of the male biased
224 *Stomoxys* genes in other species in which seminal proteins are characterized including *G.*
225 *morsitans* [45], *D. melanogaster* [46, 47], *Ae. aegypti* [48] and *Homo sapiens* [49] (Fig. 3). In
226 general, the number of orthologous sequences identified corresponded to the basic
227 phylogenetic relationships between the species tested. However, these relationships did not
228 hold for the number of gene orthologs associated with seminal function. Reciprocal analysis with
229 *Drosophila* identified 469 1:1 orthologs of male biased *Stomoxys* genes, amounting to the
230 largest number of orthologs identified between species included in this analysis. In contrast, of
231 those 469 orthologs only 1 is associated with seminal fluid function in *D. melanogaster*.
232 Comparison with *G. morsitans* identified the second highest number of orthologous proteins
233 (387). Of those, 53 were associated with seminal function, suggesting a greater similarity in the
234 constitution of seminal secretions between *Glossina* and *Stomoxys* consistent with their closer
235 phylogenetic relationship compared to *Drosophila*. Of note, none of the *Stomoxys* male biased
236 proteins were orthologous to seminal proteins across all four species. In *Drosophila*, male-
237 biased genes evolve at a faster rate, especially those expressed in reproductive tissues, and
238 they tend to lack identifiable orthologs compared with genes expressed in an unbiased pattern
239 [50, 51]. There is evidence for this in *Musca* as well [52], and these may be due to selection
240 pressure resulting in rapid evolution of male-biased genes [51]. One gene, however, a catalase
241 (XM_013259723), was orthologous to seminal protein genes in *Aedes*, *Glossina* and *Homo*
242 *sapiens* (Fig 3A). As catalases function to reduce oxidative stress, this finding could reflect a
243 conserved mechanism that protects the sperm from oxidative damage.
244 The 763 male biased *Stomoxys* genes were annotated by BLAST and gene ontology
245 analysis and then categorized by best hit annotation. Of those genes, 216 lacked significant
246 BLAST hits or were homologous to hypothetical proteins with no functional associations. Of the
247 remaining genes that had significant hits to annotated proteins, certain categories were highly
248 expressed in terms of both the number of genes and the relative level of expression within the
249 male reproductive tract (Fig. 3B+C).
250 The top category for which function could be assigned was comprised of 16 genes with
251 chitinase activity, 12 of which are clustered on scaffolds KQ079939 (7 genes) and KQ080089 (5
252 genes). Chitinases confer anti-fungal activity in honey bee seminal secretions, preventing the
253 transfer of pathogenic spores during copulation [53]. Such protective properties would be
254 beneficial to *Stomoxys* given the high probability of exposure to fungi in the moist and microbe

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

255 rich substrates in which females oviposit. The second most highly expressed category consisted
256 of a single gene, XM_013245551, which is the most highly expressed gene in the male
257 reproductive library. While it is annotated as a GATA zinc-finger domain containing protein,
258 further analysis reveals little in the way of conserved domains to indicate its function. The high
259 level of expression of this gene suggests it is an important participant in the functions of this
260 tissue. Cytochrome P450s were an additional class of male biased reproductive tract genes.
261 Cytochrome P450 proteins are involved in a wide array of processes and are associated with
262 molecular modifications, such as hormone biosynthesis and detoxification of xenobiotics [54,
263 55]. This analysis provides some insight into genetic associations with male reproductive
264 functions in *Stomoxys* and highlights a number of interesting targets for functional analysis in
265 the future. Such analyses could provide key targets for *Stomoxys* control strategies including
266 sterile male production and novel reproductive inhibitors.
267

268 **Evidence of muscid-specific odorant binding proteins and odorant receptor lineages**
269

270 Chemosensory pathways rely on gene families encoding odorant binding proteins, carrier
271 proteins for lipid molecules, as well as odorant, gustatory, and ionotropic receptors that display
272 different affinities for air-borne molecules, mediating the insect's response to its environment.
273 The *Stomoxys* and *Musca* genomes encode numerous lineage-specific expansions and contain
274 signatures of many pseudogenizations/deletions of chemosensory pathway genes relative to
275 *Drosophila* (Figs. 4 – 6, Additional File 1: Section 7), consistent with the birth-and-death model
276 of evolution proposed for these gene families [56].
277

278 *Odorant Binding Protein (OBP) Gene Family*

279 The *Stomoxys* OBP family is comprised of 90 genes, more than half of which are organized
280 as tandem clusters across three scaffolds, consistent with OBP gene organization in other
281 dipteran genomes [57] (Fig. 4, Additional File 2: Table S19). Further, two lineages of OBPs
282 appear unique to *Stomoxys* and *Musca* (Fig. 4A, orange lineage). Expression of 44 *Obp*s was
283 detected in heads of both mated, adult females and males (Fig. 4B), of which 7 *Obps* and 10
284 *Obps* were highly enriched in heads of mated males and females, respectively. This expression
285 pattern may indicate a role for the genes in mediating chemosensory interactions between the
286 sexes. There is a major expansion of 31 *Stomoxys* (ScalObp11-43) and 23 *Musca*
287 (MdomObp16-38) genes related to the *DmelObp56* gene cluster (*DmelObp56a,b,d,e,h*), with 20
288 of these *Stomoxys* genes related to *DmelObp56h* (Fig. 4A, orange shade). *DmelObp56h* has a

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

289 role in male mating behavior, as gene silencing results in distinct changes in the cuticular
290 hydrocarbon profile of male *Drosophila* and in reduction of 5-T, a hydrocarbon that is produced
291 by males and thought to delay onset of courtship [103]. While *DmelObp56h* is expressed
292 exclusively in adult *Drosophila* female and male heads, *Stomoxys* transcripts in this expansion
293 have a diverse expression profile and are detected in not only head tissue of mated males and
294 females, but also in the reproductive tract (RT) tissue of mated males, in larvae, and in mated,
295 whole adult females. Whether these genes have roles in muscid mating behavior is unknown.
296 Twenty-one *Obps* were detected by RNASeq in reproductive tract (RT) tissues of mated,
297 adult males and females (Fig. 4). This was not unexpected given expression of *Obp*s in non-
298 sensory tissues and deduced roles not related to chemosensation that are reported in other
299 dipteran species [58, 59]. The transfer of OBPs from males to females in seminal fluid occurs in
300 *Drosophila*, *Glossina*, and *Ae. aegypti* [47, 60, 61], and this may account for *Obp* enrichment in
301 *Stomoxys* male RT. *ScObp 12* and *22* are highly enriched in the RT tissue of mated females,
302 suggesting putative roles for these genes in female reproduction (Fig. 4).
303
304 *Odorant Receptor (OR) Gene Family*
305 The *Stomoxys* OR family is comprised of 74 gene models, including a lineage of four and
306 three ORs that appears unique to *Stomoxys* and *Musca*, respectively (Additional File 1, Fig. S7;
307 Additional File 2: Table S21). Seventeen *Or*s were enriched in the RT of mated males, and
308 these may have a role in sperm activation, as proposed for *An. gambiae* [62]. During stable fly
309 mating, males 'perch and dart' towards females after visual and contact pheromone recognition
310 [27]. If sperm transfer is successful, females seldom re-mate [63, 64] suggesting a shift in post-
311 mating behavior. Interestingly, three *Or*s (*ScalOr22, 54,* and *55*) were highly enriched in the RT
312 of mated females relative to all other tissues (Additional File 1, Fig. S7b), suggesting these *Or*s
313 may have a role in female reproduction, such as perceiving male pheromones transferred
314 during copulation.
315 As in *Musca* and *Glossina* sp., there is an expansion of *Stomoxys* OR genes related to
316 *DmelOr67d* and *DmelOr45a* [33, 52], and smaller subsets of *Stomoxys* ORs duplicated relative
317 to *Drosophila* are present in the *Stomoxys* genome (Additional File 1: Section 7). The ScalOr
318 expansion related to *DmelOr67d* is present with five intact genes (*ScalOr50, ScalOr54-57*) and
319 three pseudogenes (*ScalOr51-53*) organized across six scaffolds; *ScalOr54 and 55* were highly
320 enriched in the RT of mated *Stomoxys* females. In *Drosophila*, *DmelOr67d* has a role in
321 recognizing a male-specific mating pheromone, cis-vaccenyl acetate [65], that regulates mating
322 behaviors [66]. Seven ScalOr genes (*ScalOr22 – 28*) are related to *DmelOr45a*, which is

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

323 expressed solely during the larval stage in *Drosophila* and mediates the response to octyl
324 acetate, a repellent substance that induces an escape response in larvae [67]. Expression of
325 the *Scal* orthologs, however, is not restricted to the larval stage, suggesting an expanded role
326 for this receptor in *Stomoxys* adult responses.
327 Eleven *Ors* were detected in third instar larvae, none of which were highly enriched during
328 this immature lifestage. Further evaluation by non-quantitative RT-PCR detected an additional 9
329 *Or*s expressed in first and second instar but not in third instar larvae (Additional File 2: Table
330 S21). This suggested that stable flies differentially utilize odorant receptors throughout immature
331 development. Expression of all 20 of these *Ors* was not exclusive to the larval stages, and the
332 absence of larval-specific receptors in the stable fly may be a result of exposure to related
333 compounds during the immature and adult stages, e.g. host dung, detritus. This is in contrast
334 with mosquito species that occupy a larval aquatic habitat distinct from that of the adult.
335 Similarly, both female and male stable flies are blood feeders and sex-biased receptors to
336 enhance host or nutrient localization in one gender over the other may be less critical.
337
338 **Gustatory and ionotropic receptor gene family expansions support importance of bitter**
339 **taste perception**
340
341 *Gustatory Receptor (GR) Gene Family*
342 The gustatory receptor family is the more ancient of the two families that make up the insect
343 chemoreceptor superfamily [68-71], and comprises several highly divergent lineages, most
344 involved in taste but some in olfaction [72]. The OR family arose from a GR lineage at the base
345 of the Insecta [73, 74]. While the family is generally divided into three major and divergent
346 subfamilies (sugar or sweet receptors, the carbon dioxide receptors, and the bitter taste
347 receptors), a lineage within the bitter taste receptor clade has evolved into an important receptor
348 for fructose [75], and there are others involved in courtship [76].
349 The carbon dioxide, sugar, and fructose receptors are relatively well conserved in *Stomoxys*
350 and *Musca*, as is the case for many other insects. However, the bitter taste receptors reveal
351 considerable gene family evolution both with respect to the available relatives of these muscid
352 flies (*Drosophila* and *Ceratitis*) and between these two muscids. For example, a major
353 expansion (*ScalGr29-57*) encoding 49 candidate bitter taste receptors occurs in *Stomoxys*,
354 comparable to a similarly complicated set in *Musca* (*MdomGr43-64*, encoding 35 proteins).
355 Together, these form four major expanded clades in the muscids (Clade A – D, Fig. 5). Insect
356 carbon dioxide receptors are comprised of heterodimerized GRs encoded by highly conserved

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

357 gene lineages, Gr1 – 3 [77]. *Stomoxys* has the same set of carbon dioxide receptors as *Musca*,
358 Gr1 (*DmelGr21a*) and Gr3 (*DmelGr63a*), with a duplication of the Gr1 lineages present in
359 inverted orientation like *Musca*. The absence of the Gr2 lineage, which is also absent from
360 *Drosophila*, helps confirm that this loss occurred before the Muscidae and Drosophilidae split,
361 but after they separated from the Tephritidae because the Gr2 lineage is present in *Ceratitis*
362 [34].

363 The expression of 35 *Gr*s was detected in heads of mated females and males, of which 6
364 and 5 *Grs* were highly enriched in the female and male tissue, respectively. These 11 were
365 candidate bitter taste receptors, predominantly members of the expanded clades A, C, and D
366 (Additional File 1, Fig. S3b; Additional File 2, Table S22). Interestingly, 27 *Gr*s were enriched in
367 the male RT tissues (Additional File 1, Fig. S3b; Additional File 2, Table S20). Evidence from
368 *Drosophila* supports the expression of GRs in neurons that innervate testes and oviducts [78],
369 suggesting that these *Stomoxys* GRs may have a role in mediating reproductive system
370 function. Twenty-three *Grs* were detected in larvae, all of which were candidate bitter taste
371 receptors, suggesting they mediate larval bitter sensing. Determination of the ligand specificities
372 of these muscid receptors is required to fully understand the ecological significance of the
373 differential expansions and contractions of their bitter taste abilities.

374 Stable flies rely on chemosensory input to mediate localization of nutritional resources, such
375 as volatiles emitted by cattle [19, 20, 79-82] and volatiles/tastants produced from plant products
376 [25, 83, 84]. Stable fly adults likely respond to bacterial communities occupying various
377 substrates when identifying ideal ovipositional sites [12, 19, 82, 85], processing microbial
378 volatiles and assessing substrate suitability for moisture level and temperature.

379

380 *Ionotropic Receptor (IR) Gene Family*

381 The ionotropic receptor family is a variant lineage of the ancient ionotropic glutamate
382 receptor family [68, 86-88]. Like the GRs they are involved in both olfaction and gustation, as
383 well as sensing light, temperature, and humidity [88]. In *Drosophila* and most other insects
384 examined to date, Ir8a and 25a, both of which function as co-receptors with other IRs [88], are
385 highly conserved both in sequence and length and in being phylogenetically most closely
386 related to the ionotropic glutamate receptors from which this variant ionotropic receptor family of
387 chemoreceptors evolved [86, 87]. While *Stomoxys* has the expected single conserved ortholog
388 of Ir8a, surprisingly it has four paralogs of Ir25a (*ScalIr25a1-4*), the functions of which are
389 enigmatic, as such duplications of Ir25a have seldom been observed in other insects (Fig. 6).
390 The *Ir7a-g* and *11a* genes in *Drosophila* are expressed in larval and adult gustatory organs [87],

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

391 but ligands for these receptors are unknown. This subfamily is considerably expanded in both
392 *Musca* and *Stomoxys* and, given their complexities, they are not named for their *Drosophila*
393 relatives. Rather, these are part of the numbered series from Ir101, in the case of *Stomoxys* to
394 Ir121 and in *Musca* to Ir126 (Fig. 6). These IR gene family expansions strongly suggest an
395 expanded gustatory capacity.

396 A large clade of "divergent" IRs in *Drosophila* is involved in gustation and is known as the
397 Ir20a subfamily of 33 proteins [89-91]. This clade of mostly intronless genes is considerably
398 expanded in *Musca* to 53 members (*MdomIr127-179*), and even more so in *Stomoxys* to 96
399 members (*ScalIr122-217*). This subfamily consists mostly of minor expansions in *Drosophila*
400 and major expansions in the two muscids, labeled clades A-G in the tree (Fig. 6). Of note is
401 clade B, which has just one *Musca* gene (*MdomIr149*), but 15 *Stomoxys* genes (*ScalGr203-*
402 *217*), apparently related to DmelIr60e and 67b/c. Information on putative IR ligands in
403 *Drosophila* are limited to carbonation sensing and specific carbohydrates [92, 93].

404 RNA-seq detected 23 *Ir*s in heads of mated females and males of which 2 and 7 were
405 enriched in the female and male tissue, respectively. Interestingly, 5 *Ir*s were detected in the
406 female RT while 56 *Ir*s were detected in the male RT with 46 highly enriched in this male tissue
407 (Additional File 1, Fig. S4b; Additional File 2, Table S22), suggesting a potential critical role in fly
408 reproduction or reproductive behaviors. Given the striking expansions and expression pattern of
409 genes within this family, further functional studies in *Stomoxys* are warranted. Characterizing
410 the gene families involved in olfaction and gustation provides a means to identify compounds
411 that are repellent or attractive to stable flies, facilitating design of behavior-based control
412 technologies.

413

414 **Expansion of the long wavelength-sensitive Rh1 opsin subfamily with tuning substitution**
415 **evidence of diversified wavelength sensitivities**

416

417 *Global opsin conservation*

418 As is typical for the generally fast flying calyptrate flies, stable fly adults of both sexes are
419 equipped with large compound eyes in the lateral head and three ocelli positioned in the dorsal
420 head cuticle [94, 95]. Both, achromatic motion tracking and color-specific perception tasks begin
421 with the harvest of photons by members of the opsin class of G-protein coupled transmembrane
422 receptors which differ in their wavelength absorption optima. Our genomic survey in the stable
423 fly revealed conservation of most opsin gene subfamilies observed in *Drosophila* (Fig. 7a;
424 Additional File 1, Fig. S8 and Table S24). This included the UV sensitive opsin paralogs *Rh3*,

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

425 the blue sensitive opsin *Rh5*, several homologs of the long wavelength (LW) sensitive opsin
426 *Rh1* but a 1:1 ortholog of LW opsin *Rh6*, all of which are expressed in subsets of photoreceptors
427 in the compound eye retina [96]. In addition, we found 1:1 orthologs of the ocellus-specific opsin
428 Rh2 and the recently characterized UV-sensitive, deep brain opsin Rh7 [97] (Fig. 7a; Additional
429 File 1, Fig. S8 and Table S24). Overall, these findings are consistent with the
430 electrophysiological and positive phototactic sensitivity of *Stomoxys* to light in the UV, blue, and
431 green range of visible light [24, 98].
432 *Drosophila* and other higher Diptera including the Mediterranean fruit fly, *Ceratitis capitata*,
433 possess a second UV sensitive opsin gene *Rh4* (Fig. 7) [34, 99], which is not detected in the
434 stable fly genome. The same result was previously obtained in the tsetse fly [33]. As global
435 BLAST searches failed to detect Rh4 orthologs in any calyptrate genome, it can be concluded
436 that the Rh4 opsin subfamily was lost during early calyptrate evolution.
437

438 *An Rh1 opsin gene cluster in muscid Diptera*

439 A unique aspect of the *Stomoxys* opsin gene repertoire is the existence of six homologs of
440 the LW opsin *Rh1* (Fig. 7; Additional File 1, Table S23). Most higher Diptera sampled so far,
441 including related species like the tsetse fly [33] and the black blowfly *Phormia regina* [100],
442 possess a singleton Rh1 gene. Three Rh1 homologs, however, were detected in the *Musca*
443 draft genome [52]. Moreover, in both *Musca* and *Stomoxys*, these Rh1 homologs are closely
444 linked and anchored as a cluster by homologous flanking genes (Fig. 7; Additional File 1, Fig.
445 S9). This suggests that the Rh1 tandem gene clusters of the two species are homologous and
446 date back to an ancestral cluster in the last common ancestor of muscid Calyptratae. Consistent
447 with this, the *Stomoxys* and *Musca* Rh1 homologs formed a monophyletic unit in maximum
448 likelihood trees estimated from amino acid or nucleotide sequence alignments of dipteran Rh1
449 homologs (Additional File 1, Fig. S10). Moreover, each of the three *Musca* Rh1 homologs
450 grouped with strong support as 1:N orthologs with different members of the *Stomoxys* Rh1 gene
451 cluster (Fig. 7; Additional File 1, Fig. S10). Two of the *Stomoxys* Rh1 genes (Rh1.2.1 and
452 Rh1.2.2), however, lack *Musca* orthologs albeit rooting deeply into the muscid Rh1 gene clade
453 (Fig. 7; Additional File 1, Fig. S10). Integrating the information on gene linkage, it is possible to
454 conclude that the six Rh1 paralogs of *Stomoxys* originated by three early duplications before
455 separating from the *Musca* lineage. While the latter subsequently lost one of the two earliest
456 paralogs, the *Stomoxys* Rh1 cluster continued to expand by minimally one but possibly two
457 subsequent tandem gene duplications (Fig. 7a and Additional File 1, Fig. S9).
458

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

459 *A tuning site amino acid replacement differentiates two muscid Rh1 paralog subclusters*

460 While exceptional for other higher Diptera, tandem duplicated LW opsin gene clusters have
461 been found in mosquito and water strider species [101, 102]. In both cases, evidence of
462 functional paralog diversification has been detected in the form of amino acid changes that
463 affect opsin wavelength sensitivity, i.e. at tuning sites. Integrating data from butterflies and
464 *Drosophila*, the water strider study identified one high confidence tuning site that very likely
465 affects the blue vs green range wavelength specificity in LW opsins: site 17 based on the
466 numbering system developed for butterflies, which corresponds to residue 57 in *Drosophila* Rh1
467 [102-104]. Based on this criterion, the three oldest *Stomoxys* Rh1 gene cluster paralogs
468 preserved the blue-shifted wavelength specificity of the Rh1 singleton homologs of other
469 dipteran species (λ_{max} 480nm in *Drosophila*) given their conservation of the ancestral methionine
470 state at tuning site 17 (Fig. 7a; Additional File 1, Fig. S11). The three younger *Stomoxys* Rh1
471 paralogs, by contrast, share a leucine residue at tuning site 17, which is extremely rare across
472 insect LW opsins. In a survey of over 100 insect LW opsins, it was detected only in the
473 corresponding two Rh1 orthologs of *Musca* in addition to one more distantly related species in
474 thrips (Thysanoptera) [102]. Further, the physicochemical similarity of the tuning site 17 leucine
475 in the three youngest *Stomoxys* Rh1 paralogs to the pervasively conserved isoleucine residue
476 at tuning site 17 in the green-sensitive Rh6 opsins (λ_{max} 515nm in *Drosophila*) represents
477 compelling evidence that this shared derived replacement substitution defines a green-sensitive
478 subcluster in the *Stomoxys* Rh1 paralog group (Fig. 7a; Additional File 1, Fig. S11). Of note,
479 *Stomoxys* is also characterized by an expanded visual sensitivity in the red range [24, 98].
480 While this aspect could likewise be related to the Rh1 opsin cluster expansion, the red-
481 sensitivity of *Stomoxys* is shared with other calyptrate species, which preserved the ancestral
482 condition of a singleton Rh1 homolog (including *Glossina* and *Calliphora*) [105, 106]. This
483 suggests that the red-sensitivity peak of *Stomoxys* is mediated by accessory filter pigments
484 instead of one of the newly emerged Rh1 gene cluster paralogs.
485

486 *Dramatic transcript abundance differences between muscid Rh1 gene cluster members*

487 As expected, all *Stomoxys* opsin genes were characterized by significant transcript levels in
488 head vs other adult body regions (Additional File 1, Table S24). Moreover, the head tissue
489 derived RNA-seq data provided evidence that a single member of the Rh1 paralog clusters,
490 named Rh1.1.1.1, maintained the ancestral function of Rh1 as the major motion detection
491 specific opsin. The singleton opsin Rh1 of *Drosophila* is expressed in six motion detection-
492 specialized outer photoreceptors (R1-6) per ommatidium while opsins Rh5 and Rh6 are

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

493 differentially expressed in a single color-vision specialized photoreceptor (R8) per ommatidium
494 [96]. In the *Drosophila* modENCODE expression catalogue, this is reflected in up to 200 fold
495 higher transcript abundance of Rh1 opsin compared to Rh5 or Rh6 in the adult heads of both
496 sexes [107] (Fig. 7c; Additional File 1, Table S24). A similarly massive transcript abundance
497 was detected for the Rh1.1.1.1 homolog of both *Stomoxys* and *Musca*, while the remaining Rh1
498 cluster member genes were characterized by low to very low transcript levels (Fig. 7b;
499 Additional File 1, Table S24).

500 The dramatic expression level differences between the Rh1 cluster paralogs in *Stomoxys*
501 may reflect restrictions to smaller subsets of specialized photoreceptors or low level expression
502 throughout the retina. An attractive possibility for the former scenario is the male-specific 'love
503 spot' region [108]. In *Stomoxys*, this dorsofrontal expansion of the male eyes, which plays a role
504 in fast flight mating partner pursuit, translates into about 4 percent more ommatidia (~4,250)
505 compared to females (~4,050) [94], which may be reflected in the 1.8-fold expression level
506 difference of Rh1.1.1.1 between the adult head transcriptomes of male vs female *Stomoxys*
507 [109]. The presence of similarly enlarged male eyes in other calyptrate species (*Fannia fannia*,
508 *M. domestica*, *S. calcitrans*, *C. erythrocephala*, *Chrysomya megacephala*), however, suggests
509 that either this male-specific compound eye region was already present in the last common
510 ancestor of calyptrate flies or species are predisposed to evolving this trait [94, 110] (Additional
511 File 1, Fig. S8). The evolutionary origin of the muscid Rh1 cluster is thus not tightly correlated
512 with that of the 'love spot' region. Moreover, none of the Rh1 cluster paralogs are expressed in
513 a strictly male-specific manner (Fig. 7b; Additional File 1, Table S24). In contrast, one paralog
514 (Rh1.2.2) is characterized by 2 fold higher expression in the female head (Fig. 7b; Additional
515 File 1, Table S24). Single cell analysis of the Rh1 cluster paralog expression specificity may
516 reveal yet unknown specialized subregions in the stable fly compound eye retina.

517

518 *A shift in Stomoxys Rh5 vs Rh6 opsin transcript abundance compared to Drosophila*

519 A notable difference in the relative opsin transcript levels between *Stomoxys* and *Drosophila*
520 provides tentative evidence that some of the *Stomoxys* Rh1 gene cluster paralogs may have
521 adopted functions in the color-sensitive R8 photoreceptors. These photoreceptors express
522 either the blue sensitive Rh5 opsin or the green-sensitive Rh6 opsin in *Drosophila* [96]. In the
523 *Drosophila* modENCODE expression catalogue, Rh6 transcripts are 3-5 fold more abundant in
524 the whole head transcriptome in comparison to Rh5, consistent with the expression of Rh6 in
525 about 70% of the *Drosophila* R8 photoreceptors (Fig. 7c; Additional File 1, Table S24). In the
526 head transcriptomes of both male and female *Stomoxys*, by contrast, Rh6 transcripts are 2

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

527 orders of magnitude less abundant than Rh5 transcripts (Fig. 7b; Additional File 1, Table S24).
528 In the heads of both sexes, Rh6 opsin transcript abundance is the second lowest of all opsin
529 genes and even below that of the deep brain opsin Rh7, which is expressed in only about 20
530 pacemaker cells in *Drosophila* [97]. The lower transcript abundance of *Stomoxys* Rh6 relative to
531 *Drosophila* could reflect a partial (but pronounced) replacement of ancestral Rh6 expression in
532 R8 photoreceptors by the blue-sensitive Rh5 opsin or members of the Rh1 gene cluster. Of
533 possible significance in this context, stable flies exhibit strong positive phototaxis in response to
534 UV- and blue range light sources [22, 23, 111, 112]. Moreover, positive phototaxis to blue light
535 increases in female flies after fertilization [24]. Whether and how these visual preferences relate
536 to the high ratio of blue (Rh5) vs green-sensitive (Rh6) opsin expression also awaits resolution
537 through single cell expression and wavelength-sensitivity studies of the unexpectedly complex
538 opsin gene repertoire of the stable fly.
539

540 **Immune system gene family expansions may reflect adaptation to larval development in**
541 **microbe-rich substrates**

542 Analysis of the *Stomoxys* genome revealed extensive conservation of immune system
543 signaling pathways coupled with dramatic expansions of some gene families involved in both
544 recognition and effector functions. The insect immune system – best characterized from work in
545 the model organism *Drosophila melanogaster* – includes both cellular defenses (e.g.,
546 macrophage-like cells that phagocytose pathogenic microorganisms) and a humoral defense
547 system that results in the production of antimicrobial effector molecules [113]. The humoral
548 immune system can be divided into recognition proteins, which detect pathogenic bacteria and
549 fungi; signaling pathways, which are activated by recognition proteins and result in the
550 translocation of transcription factors to the nucleus to induce gene expression; and effectors,
551 which are (typically) secreted and ultimately act to clear infections.

552 Previous comparative work suggests that at least some parts of the immune system are
553 deeply conserved across Dipterans and indeed most insects. Genes encoding immune
554 signaling proteins, in particular, are generally preserved as single-copy orthologs across a wide
555 range of insects [33, 52, 114-120], with only rare exceptions [121]. Despite the strong
556 conservation of the basic structure of the main signaling pathways in insect immunity, there is
557 considerable evidence for variation in both the gene content and protein sequence of the
558 upstream inputs (recognition proteins) and downstream outputs (effector proteins) of the
559 immune system (e.g., [116-119, 122, 123]).

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

560 We find major components of the Toll, Imd, JAK/STAT, p38, and JNK pathways in the *S.*
561 *calcitrans* genome (Additional File 1, Table S27), largely conserved as single-copy orthologs
562 [113]. A description and full lists of putative computationally annotated and manually curated
563 immune-related genes in *S. calcitrans* is provided (Additional File 1, Section 9; Additional File 2,
564 Tables S25 and S26). These findings are consistent with previous reports for many other
565 Dipterans, and supports the conclusion that the intracellular signaling mechanisms of innate
566 immunity have been stable during the evolutionary history of Dipterans. In contrast, the gene
567 families encoding upstream recognition proteins and downstream effector proteins tend to be
568 expanded in *S. calcitrans* and *M. domestica* relative to other Dipterans (Table 1).
569 Based on Hidden Markov Model profiles and manual curation, we analyzed four canonical
570 recognition families with well-characterized immune roles and an additional eight families with
571 less well-defined roles (Table 1). For three of the four canonical pattern recognition receptor
572 families (PGRP, NIM, TEP), and four of the other families (CTL, GALE, FREP, and SRCB), the
573 *Stomoxys* genome encodes either the most or second-most after *M. domestica*, members
574 among the 5 Dipteran genomes screened (*S. calcitrans* plus *Aedes aegypti*, *D. melanogaster*, *.*
575 *domestica*, and *G. morsitans*). A similar pattern holds for downstream effector proteins: the *S.*
576 *calcitrans* genome encodes either the most or second-most after *M. domestica* for attacins
577 (ATT), defensins (DEF), cecropins (CEC) and lysozymes (LYS). Not unexpectedly, three
578 classes of AMPs were originally characterized in *D. melanogaster* but are missing from the *M.*
579 *domestica* genome (Metchnikowin, Drosocin, Drosomycin) and are also not detected in the
580 *Stomoxys* genome.
581 Several of the expanded AMP gene families were found clustered on individual scaffolds,
582 possibly arising from tandem duplications (Additional File 1, Section 9). For example, the 11
583 *Stomoxys defensin* genes are located on a single scaffold (KQ079966), and these
584 phylogenetically separate into three lineages that are grouped into two regions along the
585 scaffold, i.e. one includes five *defensins* present upstream of the other that includes six
586 *defensins* (Additional File 1, Fig S15). Within the downstream cluster, three genes demonstrate
587 larval-biased expression patterns while the other three appear induced upon blood-feeding in
588 adults (Additional File 2, Table S26). In contrast, *defensin* genes in the upstream cluster are all
589 detected in newly emerged adults and are upregulated in response to blood-feeding. This
590 expression pattern is consistent with that reported for *Stomoxys midgut defensin 1* (*Smd1*) and
591 *Smd2*, which were present in this upstream cluster [124].
592 The *Stomoxys* genome encodes 17 PGRPs, 11 in the short subfamily, 5 in the long
593 subfamily, and 1 ambiguous. Of the members of the short subfamily, six – all orthologs of the

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

594 PGRP-SC gene family in *D. melanogaster* – appear to be expressed exclusively in larvae and,
595 based on sequence properties and conservation of residues required for amidase activity [125],
596 are predicted to be both secreted and catalytic (Additional File 1, Fig. S13). PGRP-SC genes in
597 both *D. melanogaster* [126] and *M. domestica* [127] are also expressed in larvae, but this
598 expression is not exclusive suggesting the possibility that larval-specific expression may be an
599 *S. calcitrans* innovation.
600 In combination with the previously reported expansions of many effector and recognition
601 immune components in the house fly [52, 116], our analysis of the *Stomoxys* genome suggests
602 that Muscidae likely have expanded the diversity of their immune repertoires, sometimes
603 dramatically, despite differences in adult feeding ecology (blood feeder vs generalist). One
604 hypothesis is that the shared diversification of immune receptors and effectors is driven by larval
605 ecology (e.g., the shared requirement for bacteria during development and the septic
606 environment larvae inhabit), while additional *M. domestica* specific expansions (e.g. in TEPs)
607 are accounted for by the saprophytic adult feeding behavior of that species.
608
609 **Immunomodulatory and anti-hemostatic products are prominent in the sialome**
610 Blood-feeding insects salivate while probing their host skin for a blood meal. Development of
611 a sophisticated salivary potion that disarms their hosts' hemostasis is among the adaptations to
612 blood feeding found in hematophagous animals [128, 129]. Blood clotting inhibitors, anti-platelet
613 compounds, vasodilators and immunomodulators are found in salivary gland homogenates or
614 saliva of blood sucking arthropods [129]. To determine the genes associated with salivation, we
615 mapped the reads from four RNA-seq libraries (male and female salivary glands - SG, as well
616 as male and female whole bodies – WB) to the *S. calcitrans* predicted gene set. We used an X^2
617 test to identify those that were significantly over-expressed in SG (Fig. 8), as in [130] (Additional
618 File 2, Tables S2 and S29). A subset of SG transcripts with 100-fold higher expression than in
619 WB was analyzed in more detail (Additional File 2, Table S30). The SG 100-fold overexpressed
620 set was comprised of 139 transcripts, 18 of which were found to be splice variants, or identical
621 to other transcripts, as verified by their scaffold coordinates. The non-redundant set comprised
622 of 121 transcripts was classified into three major groups: Putative Secreted, Putative
623 Housekeeping and Unknown; these groups were further classified into finer functional
624 categories (Additional File 2: Table S29).
625 In congruence with the SDS gel of *S. calcitrans* SG [131], the antigen 5 family returned 62%
626 of the total reads mapped to the 121 SG-enriched transcripts (Fig. 8). Members of this family in
627 *S. calcitrans* may function as inhibitors of the classical complement system [132].

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

628 Thrombostasin [133] members are represented by two transcripts (29% of reads with strong gel
629 bands, [131]), which are precursors for anti-thrombin peptides previously identified in *S.*
630 *calcitrans*. They accrued 29 % of the reads and are strongly represented in the gel bands. The
631 Hyp 16 family of peptides (unknown function, 4.9% of accrued reads) and one transcript
632 encoding an endonuclease (1.2% of accrued reads) were also noted. Together, these groups of
633 transcripts account for 97% of the reads that are over expressed in the salivary glands of *S.*
634 *calcitrans.*

635 There was a wide variety of other transcripts represented in the last 3% of the reads, and
636 serine proteases, nucleotide deaminase, amylase, phospholipase A2 and lipases were found
637 enriched in the *S. calcitrans* sialome. These enzymes are also found enriched in other sialomes
638 and their functions have been reviewed [129, 134, 135]. Two of eight serine proteases were
639 found 5-15 times overexpressed in female salivary glands when compared to male glands.
640 These two products produce best matches to vitellin-degrading proteases from *M. domestica*
641 (XM_005191887.2) and may be indeed female enriched enzymes that were hitchhiked to the
642 salivary set due to their similarities to overexpressed salivary enzymes. No other peptides were
643 found above five-fold expressed in either salivary gland gender.

644 Several antimicrobial peptides appeared enriched in the *S. calcitrans* sialome, including
645 lysozyme, attacins, defensins, diptericin, a GGY rich peptide and sarcotoxin. Of these, only the
646 GGY peptide and diptericin were identified in the Sanger-based sialome description [131].
647 These peptides may help to control microbial growth in the ingested blood. Regarding
648 polypeptides with anti-proteolytic activity, in addition to thrombostasin precursors discussed
649 above, two CDS coding for serpins (however, with very low expression) and one coding for a
650 Kazal domain-containing peptide (accruing 0.3 % of the reads) were identified. While serpins
651 may modulate clotting and inflammation-related proteases, the Kazal domain peptide may be
652 related in function to vasotab, a vasodilatory peptide from a tabanid fly [136].

653 Finally, 24 transcripts accruing 0.19 % of the reads could not be functionally classified and
654 were thus assigned to the "unknown" class. These include membrane proteins
655 (XP_013114823.1 and XP_013117270.1) that are over one thousand fold over expressed in the
656 salivary glands (the first of which was identified in the Sanger sialotranscriptome) and are
657 attractive targets for gene disruption experiments to elucidate the contribution of these proteins
658 to the salivary function of *S. calcitrans*.

659

660 **Expanded cytochrome P450 gene family may enhance metabolic detoxification of**
661 **insecticides**

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

662 The family of *Stomoxys* cytochrome P450s (CYPs) identified from the current genome
663 assembly represent a substantial expansion relative to other sequenced dipteran genomes.
664 Arthropod CYPs have diverse roles in insect physiology, including ecdysteroid biosynthesis and
665 xenobiotic detoxification [137, 138]. The CYP gene family size varies among insects with
666 dipterans having large arrays, i.e. 145 in *Musca*, 86 in *Drosophila*, and 77 in *Glossina*. The 214
667 *Stomoxys* CYPs that were identified encode representatives from each of the CYP clans that
668 are typically found in insects, *i.e.* mitochondrial, CYP2, CYP3, and CYP4 (Fig. 9; Additional File
669 2, Table S31). As in *Musca*, expansions in *Stomoxys* were primarily in clans 3 and 4. The CYP4
670 clan (62 genes) was represented by the CYP4 (51 genes) family, while the CYP3 clan (107
671 genes) comprised the largest expansion of CYPs in *Stomoxys*, predominated by the CYP6 (81
672 genes) and CYP9 (16 genes) families. Together, members of the CYP4 and CYP6 families
673 represent 62% of the *Stomoxys* CytP450s, which is comparable to what was observed in *Musca*
674 [52]. Upregulation of genes in the CYP4, CYP6, and CYP9 families has been associated with
675 resistance to spinosad and pyrethroid insecticides in *Musca*, *Anopheles*, and *Drosophila* [139-
676 141], but this has yet to be investigated in *Stomoxys*. Interestingly, tandemly duplicated
677 arrangements ("blooms") of 11 CYP4D (scaffold KQ080140), 18 CYP6A (scaffold KQ080692),
678 and 16 CYP9F (scaffold KQ080085) genes are present in the *Stomoxys* genome. Given that
679 *Stomoxys* inhabits conventional livestock production settings that utilize chemical fly control
680 measures, the initial gene duplication that eventually lead to these expanded clusters may have
681 been favored because of environmental exposure to xenobiotic pressure [142].
682 Metabolic detoxification of insecticides can be further mediated by the carboxylesterase and
683 glutathione-S-transferase gene families, members of which were identified from the *Stomoxys*
684 genome assembly and were comparable to the gene families from *Drosophila* and *Musca*
685 (Additional File 1, Section 10, Figs. S16, S17, S19). Further, 26 members of the Cys-loop gated
686 ion channel (CysLGIC) superfamily, targets for several classes of insecticides [143, 144], were
687 identified in *Stomoxys* (Additional File 1, Section 10, Figure S20).
688 Stable fly population reduction relies on integrating cultural control practices, trap
689 deployment, and application of insecticides, although the latter is considered less effective for
690 control of adult populations given that stable flies do not spend as much time on their
691 mammalian host and can disperse across long distances. Interactions of stable fly adults with
692 crops and various structures to which insecticides have been applied provides avenues for
693 exposure, and use of insecticide fogging in outbreak situations contributes to this exposure.
694 There are limited studies evaluating levels of stable fly insecticide resistance in field
695 populations, but anecdotal reports of product failure for control of stable fly populations in Brazil

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

696 (Thaddeu Barros, pers. comm.) and Costa Rica (Arturo Solorzano, pers. comm.) underscores
697 its importance. Pyrethroid resistance was reported in Europe and the US [145-147], and the
698 phenotype was attributed, in part, to target site insensitivity at the *knockdown resistance* (*kdr*)
699 locus of the *voltage-sensitive sodium channel* gene [148, 149]; there is evidence indicating
700 additional mechanisms are involved. Metabolic detoxification of insecticides may contribute to
701 pyrethroid resistance in stable flies, especially given the robust expansion of the CytP450 gene
702 family. Descriptions of these insecticide resistance (ion channel and enzymatic) gene families
703 facilitates further functional studies to define these mechanisms in stable fly populations.
704

705 **Microbiota and Lateral Gene Transfers**

706 Similar to mosquitoes, stable flies will consume blood and nectar for nourishment [25, 83,
707 150]. This is different to the closely-related tsetse flies, which are obligate blood feeders. Due to
708 this limited food source, tsetse flies harbor an obligate symbiont, *Wigglesworthia glossinidae*,
709 that provides B vitamins that are present at low levels in blood [33, 151, 152]). DNA sequencing,
710 along with analysis of the assembled *Stomoxys* genome, did not reveal a distinct microbial
711 symbiont. A preliminary study of culturable bacteria among adult *Stomoxys* collected from four
712 US (Texas) dairies revealed that there are a variety of bacteria associated with adult stable flies
713 (Fig. 10; Additional File 1, Section 11; Additional File 2, Table S30). Among those cultured,
714 *Aeromonas* was the most prevalent. *Aeromonas* sp. are frequently found in aquatic
715 environments, such as irrigation water, and have been cultured from arthropods [153]. Similar to
716 mosquitoes, harbored bacterial communities are likely acquired during the larval stage or as
717 adults during ingestion of nectar or water sources, as strict blood feeders usually have reduced
718 gut microbiota [154-156]. In addition, the *Stomoxys* bacterial communities have similarity to
719 those isolated from other flies associated with filth and decomposition (e.g. blow flies and other
720 related species, [157, 158]), suggesting these bacterial components may be retained from larval
721 development or acquired by adults when visiting sites for oviposition. Of particular interest,
722 multiple potential pathogens were cultured, such as *Staphylococcus* sp. and *Bacillus* sp.
723 associated with bovine mastitis and *Vibrio* sp., suggesting the potential of this fly to act as a
724 reservoir for pathogens. Unlike bacterial community surveys of stable flies from dairies in Brazil,
725 *Salmonella* and *Escherichia coli* were not prevalent in this US sampling [159, 160].
726 A pipeline was utilized for detecting bacterial to insect lateral gene transfers (LGT), and three
727 candidate LGTs were detected (Additional File 1, Section 12). All three are derived from
728 *Wolbachia*, a common endosymbiont found in arthropods [161] that infects 40-60 percent of
729 insect species [162, 163]. While they are a common source of LGTs, likely due to their

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

730 association with the germline of their insect hosts, the *Stomoxys* strain used for the genome
731 sequencing is not infected with *Wolbachia* and there are no reports of natural occurrence of
732 *Wolbachia* in *Stomoxys* populations. Presence of these LGTs, then, may be evidence of
733 incomplete infection of the species or to LGT events from a past *Wolbachia* infection, that has
734 been subsequently lost in the species. There is no evidence that any of the three LGTs have
735 evolved into functional protein coding genes in *Stomoxys*, although one has detectable
736 expression and occurs within the 3' UTR of XM_013245585, which encodes a transcription
737 factor containing a basic leucine zipper domain. Whether expression of the LGT is biologically
738 significant is unknown, and further studies of these three LGTs are warranted.
739

740 **Evidence for transcription factors with putative role in regulation of salivation and**
741 **reproduction**

742 To determine transcription factors (TFs) that might control specific gene expression profiles,
743 we first predicted TF-encoding genes by identifying putative DNA binding domains (DBDs),
744 using a previously described approach [32, 37]. These analyses resulted in 837 predicted TFs,
745 with the highest number coming from the C_2H_2 zinc finger and homeobox structural families
746 (Fig. 11), consistent with previously analyzed insect genomes [32, 36, 37].

747 We next predicted DNA binding motifs for as many of these putative TFs as possible using a
748 previously developed method [164]. In brief, the percent of identical amino acids was calculated
749 between each *Stomoxys* TF and each eukaryotic TF with a known motif, with values exceeding
750 a TF family-specific threshold resulting in "inferred" motifs for the *Stomoxys* TFs. For example,
751 the DBD of the uncharacterized XP_013101333 protein is 92.3% identical to the DBD of the
752 *Drosophila melanogaster* gene *cropped* (FBgn0001994). Since the DNA binding motif of
753 cropped has already been experimentally determined, and the cutoff for the bHLH family of TFs
754 is 60%, we can infer that XP_013101333 will have the same binding motif as cropped. This
755 procedure resulted in inferred motifs for 285 of the *Stomoxys* TFs (34%).

756 We then performed TF binding site motif enrichment using promoter regions for groups of
757 genes with similar gene expression patterns in our RNA-seq experiments (see Methods).
758 Promoters were defined as either 500 or 2000 bp upstream of the predicted transcription start
759 site for each gene. We restricted the search to gene set/motif pairs with significant enrichment,
760 and further filtered gene set/motif pairs to cases where (1) the given motif was present in at
761 least 60% of the promoters of the gene set, (2) the given motif was present in less than 20% of
762 all gene promoters, and (3) the difference between the presence of the motif in the gene set and

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

763 promoters of all genes exceeded 40% (see Methods). Lastly, expression of each TF was
764 verified in specific tissues using our RNA-seq datasets.
765 Based on these criteria and comparative analyses between samples, seven and nine TFs,
766 respectively, were enriched in SG for the 2000 bp and 500 bp promoter regions (Fig. 11). Based
767 on the 500 bp promoter regions, two specific TFs, *proboscipedia* (XM_013251179) and
768 *orthopedia* (XM_013261230), likely regulate SG-based transcript expression. These two TFs
769 have been associated with head and salivary development in *Drosophila* [165, 166], and the
770 increased binding sites and specific expression profile suggest a role in saliva production.
771 Male- and female-enriched analysis based on stage and tissue specific RNA-seq analyses
772 identified TF targets in each of the 500 bp and 2000 bp promoter regions (Fig. 11). The four
773 most likely TFs associated with female specific genes are XM_013251807.1 (*iroquois-class*
774 *homeodomain protein*) and XM_013252765.1 (*Unc4 homeodomain protein*) based on the 500
775 bp promoter region and two other likely homeodomain proteins, XM_013245879.1
776 (uncharacterized) and XM_013261334.1 (uncharacterized), in the 2000 bp regulatory region.
777 The latter have high expression in females and female reproductive tract (Fig. 11). Two TFs
778 within the 500 bp promoter region were for male enriched genes XM_013258869.1 (*BarH-2*)
779 and XM_013251073.1 (uncharacterized), both of which are highly expressed in males, male
780 heads, and/or the male reproductive system. When expanded to the 2000 bp promoter region,
781 two additional putative TFs related to male enriched genes are XM_013260987.1
782 (uncharacterized) and XM_013257948.1 (drop), which are both highly expressed in male
783 samples. The identification of these TFs could provide novel targets for the control of stable fly
784 reproduction or the prevention of feeding.
785
786 **CONCLUSIONS**
787
788 Our analyses reveal unique aspects related to stable fly biology, including molecular
789 mechanisms underlying reproduction, chemical and host detection, feeding, and immune
790 responses. These combined studies provide substantial advancement in the understanding of
791 *Stomoxys* biology and provide the functional genomic resources to develop novel control
792 mechanisms for this livestock pest. Importantly, our studies advance the knowledge of stable fly
793 genomics and genetics to those of other non-model, but extremely important, dipterans such as
794 tsetse and house flies. Recognizing expanded vision associated genes and chemosensory
795 factors will inform the development of behavior modifying compounds and/or strategies, i.e.,
796 repellents, and enhance manipulation of visual attraction to improve traps for population

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

797 suppression. Unique proteins, including transcription factors, that are associated with
798 reproduction, feeding, and immunity will be ideal targets for gene editing strategies that modify
799 sex-specific and tissue-specific genes to aid in population suppression. Further, defining the
800 specific classes of genes that account for stable fly resistance to insecticides will enable the
801 design of diagnostic tools to monitor insecticide resistance in field populations.
802

803 **METHODS**

804

805 **Genome Sequencing, Assembly, and Annotation**

806 Multiple male (F7 generation from inbred line 8C7A2A5H3J4) individual DNA isolates were
807 provided as a pool in TE buffer. The sequencing plan followed the recommendations provided in
808 the ALLPATHS-LG assembler manual [167]. This model requires 45x sequence coverage each
809 of fragments (overlapping paired reads ~ 180bp length) and 3kb paired end (PE) reads as well
810 as 5x coverage of 8kb PE reads. For fragments and all jumping libraries (3 and 8kb) we used a
811 DNA sample pooled from approximately 10 male individuals. Total assembled sequence
812 coverage of Illumina instrument reads was 66X (overlapping reads 39x, 2kb PE 24.5x, 6kb PE
813 2.5x) using a genome size estimate of 900Mb reported by the ALLPATHS-LG software (Broad
814 Institute). This first draft assembly was referred to as S_calcitrans 1.0. In the S_calcitrans 1.0
815 assembly small scaffold gaps were closed with Illumina read mapping and local assembly, and
816 scaffolding was improved using SSPACE [168]. Contaminating contigs, trimmed vector in the
817 form of X's and ambiguous bases as N's in the sequence were removed. NCBI requires that all
818 contigs 200bp and smaller be removed. Removing these contigs was the final step in
819 preparation for submitting the 1.0.1 assembly. The final *S. calcitrans*-1.0.1 assembly is made up
820 of a total of 12,042 scaffolds with an N50 scaffold length of over 504kb (N50 contig length was
821 11kb). The total scaffold assembly including gaps and single contigs scaffolds spans over
822 971Mb. Data for the *S. calcitrans* genome have been deposited in the GenBank Bioproject
823 database under the accession code PRJNA188117. The genome assembly has been deposited
824 to GenBank under the accession GCA_001015335.1. RNA-seq datasets used in gene
825 prediction have been deposited to the NCBI Sequence Read Archive under the accession
826 codes SRX995857 - 5860, SRX229930, SX229931, and SRX275910 (Additional File 2, Table
827 S1). Methods related to the annotation are described within Additional File 1, Section 1.
828

829 **RNA-seq analyses**

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

830 In conjunction with the genomic sequencing, RNA-seq analyses were performed to examine
831 specific transcript differences between different stages and tissues (Additional File 2, Tables S5
832 – S13). RNA was extracted with the use of TRizol. DNA contamination was reduced via DNase
833 treatment according to methods previously described [169, 170]. Briefly, RNA-seq datasets
834 were trimmed and low-quality reads were removed with trimmomatic. Each dataset was mapped
835 to the to the predicted gene sets (NCBI Annotation Release 100) using CLC Genomics
836 (Qiagen). Reads were mapped with the following parameters: 95% match for over 60% of the
837 read length with only two mismatches allowed. Reads alignments were converted to per million
838 mapped to allow comparison between RNA-seq data sets with varying coverage. Expression
839 was based upon transcripts per million (TPM). A Baggerly's test followed by a Bonferroni
840 correction at 0.01 (number of genes x α value) was used to establish significance. Each RNA-
841 seq dataset was validated by qRT-PCR (Additional File 1, Section 3). Gene ontology
842 assessment for specific groups were conducted by the use of gProfiler following conversion of
843 *Stomoxys* gene IDs to *D. melanogaster* gene IDs [171]. Salivary gland RNA-seq analyses were
844 conducted based on methods used in previous studies on insect vectors [130].
845

846 **Transcription factor analyses**

847 To assess potential transcription factors regulating tissue and sex-specific expression, TFs were
848 identified according to previously developed methods in other insect systems [32, 37]. Enriched
849 TF binding motifs were identified in the 500 and 2000 bp regions upstream of the putative
850 transcription start site using the HOMER tool [172] supplemented with the *Stomoxys* inferred
851 binding motifs obtained from the CisBP database (build 0.90).
852

853 **Microbiome analyses and LGT Prediction**

854 To identify culturable bacterial communities harbored by adult stable flies, fly specimens were
855 collected at each of four Texas dairies in April and June 2015 (Lingleville and Comanche,
856 Texas). Twenty flies per site per date were collected by aerial sweep nets in the area
857 surrounding each dairy's milking parlor. Within 4 hours, whole flies were surface sterilized in 1%
858 sodium hypochlorite for 15 minutes, followed by two washes in 70% ethanol and three rinses in
859 sterile water. Individual flies were macerated in Butterfield's phosphate buffer, and the
860 homogenate was diluted and plated on tryptic soy agar. Individual, morphologically distinct
861 colonies were selected, suspended in Butterfield's phosphate buffer, and the DNA isolated by
862 rapid boiling. These DNAs were used as template in 16S PCR amplification with a universal
863 primer pair (16SEub_61F: 5' – GCTTAACACATGCAAG – 3'; 16SEub_1227R: 5' –

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

864 CCATTGTAGCACGTGT – 3'). Individual amplicons were sequenced in both directions and the
865 sequences assembled. Data were processed and full-length sequences (N=170) were analyzed
866 in mothur, v 1.38.1 [173]. Sequences were aligned using the silva.seed_v128.align file and were
867 subsequently clustered and classified at a 97% similarity cut-off, average neighbor,
868 silva.nr_v128.align/.tax.

869 A DNA based computational pipeline was used to identify "contaminating" bacterial scaffolds
870 and bacterial to *Stomoxys* lateral gene transfer (LGT) candidates in the *Stomoxys* genome
871 assembly. The pipeline was originally developed by Wheeler et al. [174] that has subsequently
872 been modified. Details of the pipeline are provided in Poynton et al. [175] and Panfilio et al.
873 [176], which have been summarized in Additional File 1, Section 12.

874

875 **Declarations**

876

877 **Ethics Approval and Consent to Participate:** Not applicable

878

879 **Consent for Publication:** Not applicable

880

881 **Availability of Data**:
882 All genome sequence data are publicly available at the NCBI BioProject: PRJNA188117, and
883 RNA-Seq transcriptome data at BioProject: PRJNA288986, with the genome assembly at NCBI
884 accession number GCA_001015335.1. RNA-seq datasets used in gene prediction have been
885 deposited to the NCBI SRA site under the accession codes SRX995857 - 5860, SRX229930,
886 SX229931, and SRX275910. Annotation and gene model data are available at the Apollo
887 instance hosted on VectorBase (https://www.vectorbase.org/organisms/stomoxys-calcitrans;
888 https://www.vectorbase.org/apollo).

889

890 **Competing Interests:** The authors declare that they have no competing interests.

891

892 **Funding:**
893 Sequencing was supported by the NIH-NHGRI grant 5U54HG00307907 to RKK. The lateral
894 gene transfer work was supported by the US National Science Foundation (DEB1257053 and
895 208 IOS1456233) to JHW. Funding for MTW was partially provided by CpG Award 53553 and a
896 CCRF Endowed Scholar Award from Cincinnati Children's Hospital Medical Center. RMW was
897 supported by Swiss National Science Foundation grant PP00P3_170664. Funding for JBB was

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

898 partially provided by bioinformatics support from the National Science Foundation (DEB-
899 1654417) and United States Department of Agriculture (CA-D-ENM-2479-CG). Funding for PUO
900 was provided by the US Department of Agriculture | Agricultural Research Service (3094-32000-
901 038-00D). The automated gene annotation work was supported by the Intramural Research
902 Program of the National Library of Medicine, National Institutes of Health.
903

904 **Footnote:**

905 This article reports the result of research only. Mention of trade names or commercial products
906 in this publication is solely for the purpose of providing specific information and does not imply
907 recommendation or endorsement by the U.S. Department of Agriculture. The USDA is an equal
908 opportunity provider and employer.
909

910 **Authors' Contributions:**

911 PUO and JBB organized and directed the sequencing, analysis, and manuscript development.
912 Grant funding was obtained for this project as part of the *Glossina* genome project by SA.
913 Sequencing and assembly of the genome was performed by WCW and RKW. Hosting of the
914 genome for annotation was provided by GLM, SJE, and DL. Gene predictions were performed
915 by TDM. Male and female reproductive analyses were conducted by GMA and JBB.
916 Transcription factor analyses were conducted by XC, ENJ, MTW, and JBB. BUSCO and
917 phylogenetic analyses were conducted by EOM, ECJ, and JBB. JMCR, JBB, and PUO collected
918 samples and performed sialome analyses. Chemosensation associated genes were annotated
919 and analyzed by HMR, MD, GB, and PUO. Cuticle proteins were analyzed by AJR and JBB.
920 Vision associated genes were analyzed by MF and JD. CJH and JBB assisted in the annotation
921 of aquaporins. Annotation of CEGMA and autophagy genes was conducted by KJO and JBB.
922 Immune genes were annotated and analyzed by TBS, DN, and PUO. DBT contributed to the
923 writing of the introduction and interpretations of the results. Muller elements were assigned and
924 analyzed by RPM. Bacterial community was assessed by SLS and PUO. Cytochrome P450s
925 were analyzed by DRN. Carboxylesterases, GSTs, and CLGICs were analyzed by PUO. JHW
926 conducted analyses of lateral gene transfers. OrthoDB analyses were conducted by EZ, RMW,
927 and PI. Transposable elements and genomic repeats were analyzed AMT, SHS, CJC, and TJR.
928 PUO, JBB, GMA, MF, RPM, DN, TJR, JMCR, HMR, TBS, AMT, DBT, MTW, RMW, and JHW
929 wrote and edited the manuscript. All authors read and approved the final manuscript.
930

931 **Acknowledgements:**

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

932 We thank The Genome Institute at Washington University for library construction and
933 sequencing. JHW thanks S Cheng for assistance with the LGT pipeline.

934

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

935 **Figures, Tables, and Additional Files**

936 **Figure 1. Quality assessment of Stomoxys genome. A**. Number of genes with alignment to
937 *Drosophila melanogaster* genome. **B**. Ortholog group comparison between *Stomoxys* and other
938 higher flies based on comparison to the OrthoDB8 database. All *Drosophila* species were
939 merged for the analysis.
940
941 **Figure 2. Maximum Likelihood Phylogenetic analysis of yolk protein genes from**
942 ***Drosophila melanogaster*, *Glossina morsitans*, *Musca domestica* and *Stomoxys***
943 ***calcitrans***. Numerical annotations indicate bootstrap values for each branch point in the tree.
944 Heat map of gene expression (transcripts per million, TPM) is based on RNA-seq data
945 (Additional File 2, Tables S5 - 13). RS, reproductive system.
946
947 **Figure 3. Analysis of male reproductive biased genes in *Stomoxys calcitrans*. A.** Results
948 of reciprocal BLAST analysis of male reproductive tract biased genes. **B.** Expression analysis of
949 top 20 most abundant gene classes as annotated by BLAST best hits. Bar length represents
950 combined expression values in RPKM for the genes included in that category. Numbers
951 associated with the bars represent the number of genes in that functional classification. **C.**
952 Scatter plot of the 763 *Stomoxys* male reproductive biased genes. The plot shows on the x axis
953 - $\log_2$ fold change expression in males relative to females and the y axis represents the $\log_2$
954 transformed expression value in RPKM in the male reproductive tissue. Triangular points are
955 genes predicted to contain signal peptides and blue points are genes with orthology to seminal
956 proteins in other species. Genes with a $\log_2$ expression value above 10 and $\log_2$ Male/Female
957 Fold Change value above 5 are annotated with putative functional descriptions.
958
959 **Figure 4. Phylogenetic tree of the *Stomoxys calcitrans* OBPs with those of *Drosophila***
960 ***melanogaster* and *Musca domestica* along with RNA-seq expression.** A. Maximum
961 likelihood phylogeny was constructed using the web server version of IQ-TREE software ([177];
962 best-fit substitution model, branch support assessed with 1000 replicates of UFBoot bootstrap
963 approximation). The *S. calcitrans* and *M. domestica* gene/protein names are highlighted in teal
964 and blue, respectively, while *D. melanogaster* names are in mustard. Clades that are expanded
965 in the muscids relative to *Drosophila* are shaded in orange, blue, yellow, and green. B. OBP
966 transcripts with detectable expression by RNASeq among tissues and developmental stages,
967 based on complete results within Additional File 2, Table S22.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

968

969 **Figure 5. Phylogenetic tree of the *Stomoxys calcitrans* GRs with those of *Drosophila*
970 *melanogaster* and *Musca domestica.*** Maximum likelihood tree rooted by declaring the
971 distantly-related and divergent carbon dioxide and sugar receptor subfamilies as the outgroup.
972 The *S. calcitrans* and *M. domestica* gene/protein names are highlighted in blue and teal,
973 respectively, while *D. melanogaster* names are in mustard. Support levels from the approximate
974 Likelihood-Ratio Test (aLRT) from PhyML v3.0 are shown on branches. Subfamilies and
975 individual or clustered *Drosophila* genes are indicated outside the circle to facilitate finding them
976 in the tree. Four clades of candidate bitter receptors that are expanded in the muscids are
977 highlighted. Pseudogenic sequences are indicated with the suffix P. Scale bar indicates amino
978 acid substitutions per site.

979

980 **Figure 6. Phylogenetic tree of the *Stomoxys calcitrans* IRs with those of *Drosophila*
981 *melanogaster* and *Musca domestica.*** Maximum likelihood tree rooted by declaring the
982 Ir8a/25a lineage as the outgroup. The *S. calcitrans* and *M. domestica* gene/protein names are
983 highlighted in blue and teal, respectively, while *D. melanogaster* names are in mustard. Support
984 levels from the approximate Likelihood-Ratio Test from PhyML v3.0 are shown on branches.
985 Subfamilies, clades, and individual *Drosophila* genes are indicated outside the circle to facilitate
986 finding them in the tree. Pseudogenic sequences are indicated with the suffix P. Scale bar
987 indicates amino acid substitutions per site.

988

989 **Figure 7. Phylogenetic and genomic organization of dipteran opsin gene relationships**. **A**.
990 Phylogenetic tree of dipteran opsin gene relationships, and genomic organization and evolution
991 of the *Stomoxys calcitrans* Rh1 opsin subfamily [178]. Protein sequences were aligned with
992 MUSCLE [179]. Ambiguous alignment regions were filtered using Gblocks [180] using least
993 stringent settings. Maximum liklihood tree was estimated in MEGA version 6.0 [181] applying
994 the Jones-Taylor-Thornton (JTT) model of amino acid sequence evolution and assuming
995 Gamma Distributed substitution rates across sites with 3 categories. **B.** Transcript abundance
996 differences between *Stomoxys* and *Drosophila* opsins.

997

998 **Figure 8. Analysis of salivary gland biased genes in *Stomoxys calcitrans.*** Number of
999 Illumina reads versus fold enrichment in the salivary gland related to the whole body. Each point
1000 represents the average among all genes in that specific category. Expression levels are based
1001 on results in Additional File 2, Table S29.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1002

1003 **Figure 9. Phylogenetic analysis of cytochrome P450 genes from *Stomoxys calcitrans*.**
1004 Amino acid sequences from each family were aligned with the MUSCLE algorithm [179], and
1005 the alignments trimmed with the trimAl tool using the –strictplus option [182]. The trimmed
1006 alignment was used to construct a maximum likelihood phylogeny, rooted with *Mus musculus*
1007 CYP51 as the outgroup, with the web server version of IQ-TREE software (best-fit substitution
1008 model, branch support assessed with 1000 replicates of UFBoot bootstrap approximation;
1009 [177]). The CYP clades are presented in different colors, and CYP gene clusters that are found
1010 in tandem within the genome are shaded in grey. P450 gene names were assigned based on
1011 comparative analyses.

1012

1013 **Figure 10. Microbiome analyses of culturable bacteria.** Survey of the culturable bacteria
1014 associated with stable flies from four separate dairies in Texas. Overlap represents species
1015 present at all four sites. Complete list is available in Additional File 2, Table S32.

1016

1017 **Figure 11. Transcription factors associated with stable flies. A**. Number of transcription
1018 factors identified in *Stomoxys calcitrans* compared to other flies. (**B** and **C**) Overlap between
1019 transcription factors with increased binding sites in differentially expressed genes that have
1020 noted expression in the same tissue (F, teneral female; FRS, female reproductive system; M,
1021 male; MRS, male reproductive system; SG, salivary glands). (**D** and **E**). Expression of specific
1022 TFs associated with female, male, and salivary glands among multiple tissues and
1023 developmental stages.

1024

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1025 **Table 1. Number of Immune-Related Gene Families Annotated by Hidden Markov Models.**
1026 Numbers in parentheses are those numbers annotated after manual curation of the genome.

		Scal	***Mdom***	***Gmor***	***Aaeg***	***Dmel***
	ATT	11 (12)*	10	4	1	4
	DEF	5 (11)*	5	0	4	1
	DIPT	3 (1)*	4	0	1	3
	CEC	5 (10)*	12	2	9	5
canonical effectors	LYS	23	32	4	7	13
	TPX	5	6	6	5	8
	PPO	19	23	4	25	10
	GPX	1	1	0	3	2
	HPX	12	12	8	19	10
non-canonical effectors	TSF	4	6	3	5	3
	NIM	25	23	10	8	17
	PGRP	17	17	4	10	13
	BGBP	4	3	3	7	7
canonical recognition	TEP	16	22	4	8	6
	CTL	78	41	11	43	38
	FREP	49	38	7	34	14
	GALE	15	13	8	12	6
	IGSF	1	1	1	0	1
	MD2L	8	12	5	26	8
	SRCA	3	3	2	2	3
	SRCB	15	18	11	13	14
other recognition	SRCC	7	8	4	5	9

1027

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1028 **References.**

1029

1030 1. Krasfur ES, Moon RD: **Bionomics of the face fly, *Musca autumnalis*.** *Annu Rev*
1031 *Entomol* 1997, **42:**503-523.
1032 2. Foil LD, Hogsette JA: **Biology and control of tabanids, stable flies and horn flies.**
1033 *Rev Sci Tech OIE* 1994, **13:**1125-1158.
1034 3. Dougherty CT, Knapp FW, Burrus PB, Willis DC, Burg JG, Cornelius PL, Bradley NW:
1035 **Stable flies (*Stomoxys calcitrans* L) and the behavior of grazing beef cattle.** *Appl*
1036 *Anim Behav Sci* 1993, **35:**215-233.
1037 4. Dougherty CT, Knapp FW, Burrus PB, Willis DC, Cornelius PL: **Moderation of grazing**
1038 **behavior of beef cattle by stable flies (*Stomoxys calcitrans* L).** *Appl Anim Behav Sci*
1039 1994, **40:**113-127.
1040 5. Dougherty CT, Knapp FW, Burrus PB, Willis DC, Cornelius PL: **Behavior of grazing**
1041 **cattle exposed to small populations of stable flies (*Stomoxys calcitrans* L).** *Appl*
1042 *Anim Behav Sci* 1995, **42:**231-248.
1043 6. Mullens BA, Lii KS, Mao Y, Meyer JA, Peterson NG, Szijj CE: **Behavioural responses**
1044 **of dairy cattle to the stable fly, *Stomoxys calcitrans*, in an open field environment.**
1045 *Med Vet Entomol* 2006, **20:**122-137.
1046 7. Taylor DB, Moon RD, Mark DR: **Economic impact of stable flies (Diptera: Muscidae)**
1047 **on dairy and beef cattle production.** *J Med Entomol* 2012, **49:**198-209.
1048 8. Cook DF, Dadour IR, Voss SC: **Management of stable fly and other nuisance flies**
1049 **breeding in rotting vegetable matter associated with horticultural crop production.**
1050 *Int J Pest Manage* 2011, **57:**315-320.
1051 9. Dominghetti TF, de Barros AT, Soares CO, Cancado PH: ***Stomoxys calcitrans***
1052 **(Diptera: Muscidae) outbreaks: current situation and future outlook with emphasis**
1053 **on Brazil.** *Rev Bras Parasitol Vet* 2015, **24:**387-395.
1054 10. Solorzano JA, Gilles J, Bravo O, Vargas C, Gomez-Bonilla Y, Bingham GV, Taylor DB:
1055 **Biology and trapping of stable flies (Diptera: Muscidae) developing in pineapple**
1056 **residues (Ananas comosus) in Costa Rica.** *J Insect Sci* 2015, **15.**
1057 11. Cook DF, Telfer DV, Lindsey JB, Deyl RA: **Substrates across horticultural and**
1058 **livestock industries that support the development of stable fly, *Stomoxys***
1059 ***calcitrans* (Diptera: Muscidae).** *Austral Entomol* 2018, **57:**344-348.
1060 12. Romero A, Broce A, Zurek L: **Role of bacteria in the oviposition behaviour and larval**
1061 **development of stable flies.** *Med Vet Entomol* 2006, **20:**115-121.
1062 13. Baldacchino F, Muenworn V, Desquesnes M, Desoli F, Charoenviriyaphap T, Duvallet
1063 G: **Transmission of pathogens by *Stomoxys* flies (Diptera, Muscidae): a review.**
1064 *Parasite* 2013, **20.**
1065 14. Boulanger N, Munks RJL, Hamilton JV, Vovelle F, Brun R, Lehane MJ, Bulet P:
1066 **Epithelial innate immunity - A novel antimicrobial peptide with antiparasitic**
1067 **activity in the blood-sucking insect *Stomoxys calcitrans*.** *J Biol Chem* 2002,
1068 **277:**49921-49926.
1069 15. Nayduch D, Burrus RG: **Flourishing in filth: House fly-microbe interactions across**
1070 **life history.** *Ann Entomol Soc Amer* 2017, **110:**6-18.
1071 16. Scully E, Friesen K, Wienhold B, Durso LM: **Microbial communities associated with**
1072 **stable fly (Diptera: Muscidae) larvae and their developmental substrates.** *Ann*
1073 *Entomol Soc Amer* 2017, **110:**61-72.
1074 17. Gatehouse AG, Lewis CT: **Host location behavior of *Stomoxys calcitrans*.** *Entomol*
1075 *Exp Appl* 1973, **16:**275-290.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1076 18. Uebel EC, Sonnet PE, Bierl BA, Miller RW: **Sex pheromone of the stable fly: Isolation**
1077 **and preliminary identification of compounds that induce mating strike behavior.** *J*
1078 *Chem Ecol* 1975, **1:**377-385.
1079 19. Jeanbourquin P, Guerin PM: **Chemostimuli implicated in selection of oviposition**
1080 **substrates by the stable fly *Stomoxys calcitrans*.** *Med Vet Entomol* 2007, **21:**209-
1081 216.
1082 20. Jeanbourquin P, Guerin PM: **Sensory and behavioural responses of the stable fly**
1083 ***Stomoxys calcitrans* to rumen volatiles.** *Med Vet Entomol* 2007, **21:**217-224.
1084 21. Gerry AC, Peterson NG, Mullens BA: **Predicting and controlling stable flies on**
1085 **California dairies.** *University of California, Division of Agriculture and Natural*
1086 *Resources* 2007, **Publication 8258**.
1087 22. Beresford DV, Sutcliffe JF: **Studies on the effectiveness of Coroplast sticky traps for**
1088 **sampling stable flies (Diptera : Muscidae), including a comparison to Alsynite.** *J*
1089 *Econ Entomol* 2006, **99:**1025-1035.
1090 23. Taylor DB, Berkebile D: **Comparative efficiency of six stable fly (Diptera : Muscidae)**
1091 **traps.** *J Econ Entomol* 2006, **99:**1415-1419.
1092 24. Zhu JWJ, Zhang QH, Taylor DB, Friesen KA: **Visual and olfactory enhancement of**
1093 **stable fly trapping.** *Pest Manag Sci* 2016, **72:**1765-1771.
1094 25. Taylor DB, Berkebile DR: **Sugar feeding in adult stable flies.** *Environ Entomol* 2008,
1095 **37:**625-629.
1096 26. Anderson JR: **Mating behavior of *Stomoxys calcitrans* (Diptera:Muscidae): Effects**
1097 **of a blood meal on mating drive of males and its necessity as a prerequisite for**
1098 **proper insemination of females.** *J Econ Entomol* 1978, **71:**379-386.
1099 27. Yuval B: **Mating systems of blood-feeding flies.** *Annu Rev Entomol* 2006, **51:**413-440.
1100 28. Patton WS: **Studies on the higher Diptera of medical and veterinary importance.**
1101 *Ann Trop Med Parasit* 1933, **27:**135-156.
1102 29. Detinova TS: *Age-grouping Methods in Diptera of Medical Importance, with Special*
1103 *Reference to Some Vectors of Malaria.* Geneva: World Health Organization; 1962.
1104 30. Thomas AW, Gooding RH: **Digestive processes of hematophagous insects. VIII.**
1105 **Estimation of meal size and demonstration of trypsin in horse flies and deer flies**
1106 **(Diptera: Tabanidae).** *J Med Entomol* 1976, **13:**131-136.
1107 31. Picard CJ, Johnston JS, Tarone AM: **Genome sizes of forensically relevant Diptera.** *J*
1108 *Med Entomol* 2012, **49:**192-197.
1109 32. Benoit JB, Adelman ZN, Reinhardt K, Dolan A, Poelchau M, Jennings EC, Szuter EM,
1110 Hagan RW, Gujar H, Shukla JN, et al: **Unique features of a global human**
1111 **ectoparasite identified through sequencing of the bed bug genome.** *Nat Commun*
1112 2016, **7**.
1113 33. International Glossina Genome Initiative: **Genome sequence of the tsetse fly**
1114 **(*Glossina morsitans*): Vector of African trypanosomiasis.** *Science* 2014, **344:**380-
1115 386.
1116 34. Papanicolaou A, Schetelig MF, Arensburger P, Atkinson PW, Benoit JB, Bourtzis K,
1117 Castanera P, Cavanaugh JP, Chao H, Childers C, et al: **The whole genome sequence**
1118 **of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into**
1119 **the biology and adaptive evolution of a highly invasive pest species.** *Genome Biol*
1120 2016, **17**.
1121 35. Parra G, Bradnam K, Ning Z, Keane T, Korf I: **Assessing the gene space in draft**
1122 **genomes.** *Nucleic Acids Res* 2009, **37:**289-297.
1123 36. Panfilio KA, Vargas Jentzsch IM, Benoit JB, Erezyilmaz D, Suzuki Y, Colella S,
1124 Robertson HM, Poelchau MF, Waterhouse RM, Ioannidis P, et al: **Molecular**
1125 **evolutionary trends and feeding ecology diversification in the Hemiptera,**
1126 **anchored by the milkweed bug genome.** *bioRxiv* 2017**:**https://doi.org/10.1101/201731.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1127 37. Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, Brevik K,
1128 Cappelle K, Chen MJM, Childers AK, et al: **A model species for agricultural pest**
1129 **genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata***
1130 **(Coleoptera: Chrysomelidae).** *Sci Rep-UK* 2018, **8**.
1131 38. Stevens L: **Egg proteins: what are their functions?** *Sci Prog* 1996, **79 (Pt 1):**65-87.
1132 39. Chen A, Kim H, Mayer R, Norman J: **Vitellogenesis in the stable fly, *Stomoxys***
1133 ***calcitrans*.** *Comp Biochem Physiol B* 1987, **88:**897-903.
1134 40. Barnett T, Pachl C, Gergen JP, Wensink PC: **The isolation and characterization of**
1135 ***Drosophila* yolk protein genes.** *Cell* 1980, **21:**729-738.
1136 41. White NM, Bownes M: **Cloning and characterization of three *Musca domestica* yolk**
1137 **protein genes.** *Insect Mol Biol* 1997, **6:**329-341.
1138 42. Martinez A, Bownes M: **The sequence and expression pattern of the *Calliphora***
1139 ***erythrocephala* yolk protein A and B genes.** *J Mol Evol* 1994, **38:**336-351.
1140 43. Hens K, Macours N, Claeys I, Francis C, Huybrechts R: **Cloning and expression of the**
1141 **yolk protein of the tsetse fly *Glossina morsitans morsitans*.** *Insect Biochem Mol Biol*
1142 2004, **34:**1281-1287.
1143 44. Tufail M, Takeda M: **Molecular characteristics of insect vitellogenins.** *J Insect*
1144 *Physiol* 2008, **54:**1447-1458.
1145 45. Scolari F, Benoit JB, Michalkova V, Aksoy E, Takac P, Abd-Alla AM, Malacrida AR,
1146 Aksoy S, Attardo GM: **The spermatophore in *Glossina morsitans morsitans*:**
1147 **Insights into male contributions to reproduction.** *Sci Rep* 2016, **6:**20334.
1148 46. Findlay GD, MacCoss MJ, Swanson WJ: **Proteomic discovery of previously**
1149 **unannotated, rapidly evolving seminal fluid genes in *Drosophila*.** *Genome Res*
1150 2009, **19:**886-896.
1151 47. Findlay GD, Yi X, Maccoss MJ, Swanson WJ: **Proteomics reveals novel *Drosophila***
1152 **seminal fluid proteins transferred at mating.** *PLoS Biol* 2008, **6:**e178.
1153 48. Sirot LK, Hardstone MC, Helinski ME, Ribeiro JM, Kimura M, Deewatthanawong P,
1154 Wolfner MF, Harrington LC: **Towards a semen proteome of the dengue vector**
1155 **mosquito: protein identification and potential functions.** *PLoS Negl Trop Dis* 2011,
1156 **5:**e989.
1157 49. Pilch B, Mann M: **Large-scale and high-confidence proteomic analysis of human**
1158 **seminal plasma.** *Genome Biol* 2006, **7:**R40.
1159 50. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L,
1160 Artieri CG, Wolfner MF, Civetta A, Singh RS: **Evolution in the fast lane: rapidly**
1161 **evolving sex-related genes in *Drosophila*.** *Genetics* 2007, **177:**1321-1335.
1162 51. Meisel RP: **Towards a more nuanced understanding of the relationship between**
1163 **sex-biased gene expression and rates of protein-coding sequence evolution.** *Mol*
1164 *Biol Evol* 2011, **28:**1893-1900.
1165 52. Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones
1166 AK, Kasai S, Leichter CA, et al: **Genome of the house fly, *Musca domestica* L., a**
1167 **global vector of diseases with adaptations to a septic environment.** *Genome Biol*
1168 2014, **15**.
1169 53. Peng Y, Grassl J, Millar AH, Baer B: **Seminal fluid of honeybees contains multiple**
1170 **mechanisms to combat infections of the sexually transmitted pathogen *Nosema***
1171 ***apis*.** *Proc Biol Sci* 2016, **283**.
1172 54. Iga M, Kataoka H: **Recent studies on insect hormone metabolic pathways mediated**
1173 **by cytochrome P450 enzymes.** *Biol Pharm Bull* 2012, **35:**838-843.
1174 55. Rewitz KF, Rybczynski R, Warren JT, Gilbert LI: **The Halloween genes code for**
1175 **cytochrome P450 enzymes mediating synthesis of the insect moulting hormone.**
1176 *Biochem Soc Trans* 2006, **34:**1256-1260.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1177 56. Vieira FG, Sanchez-Gracia A, Rozas J: **Comparative genomic analysis of the**
1178 **odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and**
1179 **birth-and-death evolution.** *Genome Biol* 2007, **8**.
1180 57. Vieira FG, Rozas J: **Comparative genomics of the odorant-binding and**
1181 **chemosensory protein gene families across the Arthropoda: Origin and**
1182 **evolutionary history of the chemosensory system.** *Genome Biol Evol* 2011, **3:**476-
1183 490.
1184 58. Brito NF, Moreira MF, Melo ACA: **A look inside odorant-binding proteins in insect**
1185 **chemoreception.** *J Insect Physiol* 2016, **95:**51-65.
1186 59. Graham LA, Davies PL: **The odorant-binding proteins of *Drosophila melanogaster*:**
1187 **annotation and characterization of a divergent gene family.** *Gene* 2002, **292:**43-55.
1188 60. Li S, Picimbon JF, Ji SD, Kan YC, Qiao CL, Zhou JJ, Pelosi P: **Multiple functions of an**
1189 **odorant-binding protein in the mosquito *Aedes aegypti*.** *Biochem Bioph Res Com*
1190 2008, **372:**464-468.
1191 61. Scolari F, Benoit JB, Michalkova V, Aksoy E, Takac P, Abd-Alla AMM, Malacrida AR,
1192 Aksoy S, Attardo GM: **The Spermatophore in Glossina morsitans morsitans:**
1193 **Insights into Male Contributions to Reproduction.** *Sci Rep-UK* 2016, **6**.
1194 62. Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel LJ: **Transcriptome profiling of**
1195 **chemosensory appendages in the malaria vector *Anopheles gambiae* reveals**
1196 **tissue- and sex-specific signatures of odor coding.** *BMC Genomics* 2011, **12**.
1197 63. Harris RL, Miller JA, Frazar ED: **Horn flies and stable flies: Feeding activity.** *Ann*
1198 *Entomol Soc Amer* 1973, **67:**891-894.
1199 64. Smittle BJ, Seawright JA, Willis NL: **Mating competitiveness of irradiated male stable**
1200 **flies, *Stomoxys calcitrans* (Diptera: Muscidae), from a genetic sexing strain.** *J Med*
1201 *Entomol* 1984, **21:**179-182.
1202 65. van der Goes van Naters W, Carlson JR: **Receptors and neurons for fly odors in**
1203 ***Drosophila*.** *Curr Biol* 2007, **17:**606-612.
1204 66. Kurtovic A, Widmer A, Dickson BJ: **A single class of olfactory neurons mediates**
1205 **behavioural responses to a *Drosophila* sex pheromone.** *Nature* 2007, **446:**542-546.
1206 67. Bellmann D, Richardt A, Freyberger R, Nuwal N, Schwarzel M, Fiala A, Stortkuhl KF:
1207 **Optogenetically induced olfactory stimulation in *Drosophila* larvae reveals the**
1208 **neuronal basis of odor-aversion behavior.** *Front Behav Neurosci* 2010, **4**.
1209 68. Benton R: **Multigene family evolution: Perspectives from insect chemoreceptors.**
1210 *Trends Ecol Evol* 2015, **30:**590-600.
1211 69. Robertson HM: **The insect chemoreceptor superfamily is ancient in animals.** *Chem*
1212 *Senses* 2015, **40:**609-614.
1213 70. Robertson HM, Warr CG, Carlson JR: **Molecular evolution of the insect**
1214 **chemoreceptor gene superfamily in *Drosophila melanogaster*.** *Proc Natl Acad Sci*
1215 *USA* 2003, **100:**14537-14542.
1216 71. Saina M, Busengdal H, Sinigaglia C, Petrone L, Oliveri P, Rentzsch F, Benton R: **A**
1217 **cnidarian homologue of an insect gustatory receptor functions in developmental**
1218 **body patterning.** *Nat Commun* 2015, **6**.
1219 72. Joseph RM, Carlson JR: ***Drosophila* chemoreceptors: A molecular interface**
1220 **between the chemical world and the brain.** *Trends Genet* 2015, **31:**683-695.
1221 73. Brand P, Robertson HM, Lin W, Pothula R, Klingeman WE, Jurat-Fuentes JL, Johnson
1222 BR: **The origin of the odorant receptor gene family in insects.** *Elife* 2018, **7:**e38340.
1223 74. Robertson HM: **Molecular evolution of the major arthropod chemoreceptor gene**
1224 **families.** *Annu Rev Entomol* 2019, **64:**1.
1225 75. Miyamoto T, Slone J, Song XY, Amrein H: **A fructose receptor functions as a nutrient**
1226 **sensor in the *Drosophila* brain.** *Cell* 2012, **151:**1113-1125.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1227 76. Fan P, Manoli DS, Ahmed OM, Chen Y, Agarwal N, Kwong S, Cai AG, Neitz J, Renslo
1228 A, Baker BS, Shah NM: **Genetic and neural mechanisms that inhibit *Drosophila***
1229 **from mating with other species.** *Cell* 2013, **154:**89-102.
1230 77. Robertson HM, Kent LB: **Evolution of the gene lineage encoding the carbon dioxide**
1231 **receptor in insects.** *J Insect Sci* 2009, **9**.
1232 78. Park JH, Kwon JY: **A systematic analysis of *Drosophila* gustatory receptor gene**
1233 **expression in abdominal neurons which project to the central nervous system.**
1234 *Mol Cells* 2011, **32:**375-381.
1235 79. Birkett MA, Agelopoulos N, Jensen KM, Jespersen JB, Pickett JA, Prijs HJ, Thomas G,
1236 Trapman JJ, Wadhams LJ, Woodcock CM: **The role of volatile semiochemicals in**
1237 **mediating host location and selection by nuisance and disease-transmitting cattle**
1238 **flies.** *Med Vet Entomol* 2004, **18:**313-322.
1239 80. Schofield S, Cork A, Brady J: **Electroantennogram responses of the stable fly,**
1240 ***Stomoxys calcitrans*, to components of host odour.** *Physiol Entomol* 1995, **20:**273-
1241 280.
1242 81. Tangtrakulwanich K, Albuquerque TA, Brewer GJ, Baxendale FP, Zurek L, Miller DN,
1243 Taylor DB, Friesen KA, Zhu JJ: **Behavioural responses of stable flies to cattle**
1244 **manure slurry associated odourants.** *Med Vet Entomol* 2015, **29:**82-87.
1245 82. Tangtrakulwanich K, Chen H, Baxendale F, Brewer G, Zhu JJ: **Characterization of**
1246 **olfactory sensilla of *Stomoxys calcitrans* and electrophysiological responses to**
1247 **odorant compounds associated with hosts and oviposition media.** *Med Vet*
1248 *Entomol* 2011, **25:**327-336.
1249 83. Jones CJ, Milne DE, Patterson RS, Schreiber ET, Milio JA: **Nectar feeding by**
1250 ***Stomoxys calcitrans* (Diptera: Muscidae): effects on reproduction and survival.**
1251 *Environ Entomol* 1992, **21:**141-147.
1252 84. Muller GC, Hogsette JA, Beier JC, Traore SF, Toure MB, Traore MM, Bah S, Doumbia
1253 S, Schlein Y: **Attraction of *Stomoxys* sp. to various fruits and flowers in Mali.** *Med*
1254 *Vet Entomol* 2012, **26:**178-187.
1255 85. Albuquerque TA, Zurek L: **Temporal changes in the bacterial community of animal**
1256 **feces and their correlation with stable fly oviposition, larval development, and**
1257 **adult fitness.** *Front Microbiol* 2014, **5:**590.
1258 86. Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB: **Variant ionotropic glutamate**
1259 **receptors as chemosensory receptors in *Drosophila*.** *Cell* 2009, **136:**149-162.
1260 87. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton
1261 R: **Ancient protostome origin of chemosensory ionotropic glutamate receptors**
1262 **and the evolution of insect taste and olfaction.** *PLoS Genet* 2010, **6:**e1001064.
1263 88. Rytz R, Croset V, Benton R: **Ionotropic Receptors (IRs): Chemosensory ionotropic**
1264 **glutamate receptors in *Drosophila* and beyond.** *Insect Biochem Mol Biol* 2013,
1265 **43:**888-897.
1266 89. Koh TW, He Z, Gorur-Shandilya S, Menuz K, Larter NK, Stewart S, Carlson JR: **The**
1267 ***Drosophila* IR20a clade of ionotropic receptors are candidate taste and**
1268 **pheromone receptors.** *Neuron* 2014, **83:**850-865.
1269 90. Stewart S, Koh TW, Ghosh AC, Carlson JR: **Candidate ionotropic taste receptors in**
1270 **the *Drosophila* larva.** *Proc Natl Acad Sci USA* 2015, **112:**4195-4201.
1271 91. Sánchez-Alcañiz JA, Silbering AF, Croset V, Zappia G, Sivasubramaniam AK, Abuin L,
1272 Sahai SY, Münch D, Steck K, Auer TO: **An expression atlas of variant ionotropic**
1273 **glutamate receptors identifies a molecular basis of carbonation sensing.** *Nat*
1274 *Commun* 2018, **9:**4252.
1275 92. Joseph RM, Sun JS, Tam E, Carlson JR: **A receptor and neuron that activate a**
1276 **circuit limiting sucrose consumption.** *Elife* 2017, **6:**e24992.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1277 93. Sanchez-Alcaniz JA, Silbering AF, Croset V, Zappia G, Sivasubramaniam AK, Abuin L,
1278 Sahai SY, Munch D, Steck K, Auer TO, et al: **An expression atlas of variant**
1279 **ionotropic glutamate receptors identifies a molecular basis of carbonation**
1280 **sensing.** *Nat Commun* 2018, **9:**4252.
1281 94. Beersma DGM, Stavenga DG, Kuiper JW: **Retinal Lattice, visual field, and**
1282 **binocularities in flies - Dependence on species and sex.** *J Comp Physiol* 1977,
1283 **119:**207-220.
1284 95. Chayka SY, Mazokhin-Porshnyakov GA: **Ultrastructure of the compound eye of the**
1285 **stable fly, *Stomoxys calcitrans.*** *Entomol Rev* 1986, **65:**148-154.
1286 96. Wernet MF, Perry MW, Desplan C: **The evolutionary diversity of insect retinal**
1287 **mosaics: common design principles and emerging molecular logic.** *Trends Genet*
1288 2015, **31:**316-328.
1289 97. Ni JD, Baik LS, Holmes TC, Montell C: **A rhodopsin in the brain functions in**
1290 **circadian photoentrainment in *Drosophila*.** *Nature* 2017, **545:**340-344.
1291 98. Ballard RC: **Responses of Stomoxys calcitrans (L.) to radiant energy and their**
1292 **relation to absorption charcteristics of the eye.** *Ann Entomol Soc Amer* 1958,
1293 **51:**449-464.
1294 99. Fortini ME, Rubin GM: **Analysis of cis-acting requirements of the Rh3 and Rh4**
1295 **genes reveals a bipartite organization to rhodopsin promoters in *Drosophila***
1296 ***melanogaster*.** *Gene Dev* 1990, **4:**444-463.
1297 100. Andere AA, Ii RNP, Ray DA, Picard CJ: **Genome sequence of *Phormia regina* Meigen**
1298 **(Diptera: Calliphoridae): implications for medical, veterinary and forensic**
1299 **research.** *BMC Genomics* 2016, **17.**
1300 101. Giraldo-Calderon GI, Zanis MJ, Hill CA: **Retention of duplicated long-wavelength**
1301 **opsins in mosquito lineages by positive selection and differential expression.**
1302 *BMC Evol Biol* 2017, **17:**84.
1303 102. Armisen D, Rajakumar R, Friedrich M, Benoit JB, Robertson HM, Panfilio KA, Ahn SJ,
1304 Poelchau MF, Chao H, Dinh H, et al: **The genome of the water strider Gerris buenoi**
1305 **reveals expansions of gene repertoires associated with adaptations to life on the**
1306 **water.** *BMC Genomics* 2018, **19:**832.
1307 103. Frentiu FD, Bernard GD, Cuevas CI, Sison-Mangus MP, Prudic KL, Briscoe AD:
1308 **Adaptive evolution of color vision as seen through the eyes of butterflies.** *Proc*
1309 *Natl Acad Sci U S A* 2007, **104 Suppl 1:**8634-8640.
1310 104. Frentiu FD, Bernard GD, Sison-Mangus MP, Brower AV, Briscoe AD: **Gene duplication**
1311 **is an evolutionary mechanism for expanding spectral diversity in the long-**
1312 **wavelength photopigments of butterflies.** *Mol Biol Evol* 2007, **24:**2016-2028.
1313 105. Green CH, Cosens D: **Spectral responses of the tsetse fly, *Glossina morsitans***
1314 ***morsitans*.** *J Insect Physiol* 1983, **29:**795-800.
1315 106. Walther JB, Dodt E: **Die spektralsensitivitat von insekten-komplexaugen im**
1316 **ultraviolett bis 290 M-Mu - elektrophysiologische messungen an *Calliphora* und**
1317 ***Periplaneta*.** *Z Naturforsch Pt B* 1959, **14:**273-278.
1318 107. Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, Suzuki AM, Fletcher AR, Plachetzki
1319 DC, FitzGerald PC, et al: **Comparative validation of the D. melanogaster**
1320 **modENCODE transcriptome annotation.** *Genome Res* 2014, **24:**1209-1223.
1321 108. Perry MW, Desplan C: **Love spots.** *Curr Biol* 2016, **26:**R484-R485.
1322 109. Masmeatathip R, Ketavan C, Duvallet G: **Morphological studies of *Stomoxys* spp.**
1323 **(Diptera: Muscidae) in central Thailand.** *Kasetsart J (Nat Sci)* 2006, **40:**872-881.
1324 110. Smith JL, Palermo NA, Theobald JC, Wells JD: **Body size, rather than male eye**
1325 **allometry, explains *Chrysomya megacephala* (Diptera: Calliphoridae) activity in**
1326 **low light.** *J Insect Sci* 2015, **15.**

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1327 111. Allan SA, Day JF, Edman JD: **Visual ecology of biting flies.** *Annu Rev Entomol* 1987,
1328 **32:**297-316.
1329 112. Waldbillig RC: **Color vision of the female stable fly, *Stomoxys calcitrans*.** *Ann*
1330 *Entomol Soc Amer* 1968, **61:**789-791.
1331 113. Buchon N, Silverman N, Cherry S: **Immunity in *Drosophila melanogaster* - from**
1332 **microbial recognition to whole-organism physiology.** *Nat Rev Immunol* 2014,
1333 **14:**796-810.
1334 114. Barribeau SM, Sadd BM, du Plessis L, Brown MJF, Buechel SD, Cappelle K, Carolan
1335 JC, Christiaens O, Colgan TJ, Erler S, et al: **A depauperate immune repertoire**
1336 **precedes evolution of sociality in bees.** *Genome Biol* 2015, **16.**
1337 115. Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ,
1338 Zou Z, Hultmark D: **Immune pathways and defence mechanisms in honey bees *Apis***
1339 ***mellifera*.** *Insect Mol Biol* 2006, **15:**645-656.
1340 116. Sackton TB, Lazzaro BP, Clark AG: **Rapid expansion of immune-related gene**
1341 **families in the house fly, *Musca domestica*.** *Mol Biol Evol* 2017, **34:**857-872.
1342 117. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG: **Dynamic**
1343 **evolution of the innate immune system in *Drosophila*.** *Nat Genet* 2007, **39:**1461-
1344 1468.
1345 118. Sackton TB, Werren JH, Clark AG: **Characterizing the infection-induced**
1346 **transcriptome of *Nasonia vitripennis* reveals a preponderance of taxonomically-**
1347 **restricted immune genes.** *PLoS One* 2013, **8.**
1348 119. Waterhouse RM, Kriventseva EV, Meister S, Xi ZY, Alvarez KS, Bartholomay LC,
1349 Barillas-Mury C, Bian GW, Blandin S, Christensen BM, et al: **Evolutionary dynamics of**
1350 **immune-related genes and pathways in disease-vector mosquitoes.** *Science* 2007,
1351 **316:**1738-1743.
1352 120. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Beukeboom
1353 LW, Desplan C, Elsik CG, Grimmelikhuijzen CJP, et al: **Functional and evolutionary**
1354 **insights from the genomes of three parasitoid *Nasonia* species.** *Science* 2010,
1355 **327:**343-348.
1356 121. Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, De Vos M, Duncan EJ,
1357 Evans JD, Gabaldon T, Ghanim M, et al: **Immunity and other defenses in pea aphids,**
1358 ***Acyrthosiphon pisum*.** *Genome Biol* 2010, **11.**
1359 122. Hanson MA, Hamilton PT, Perlman SJ: **Immune genes and divergent antimicrobial**
1360 **peptides in flies of the subgenus *Drosophila*.** *BMC Evol Biol* 2016, **16.**
1361 123. Obbard DJ, Welch JJ, Kim KW, Jiggins FM: **Quantifying adaptive evolution in the**
1362 ***Drosophila* immune system.** *PLoS Genet* 2009, **5.**
1363 124. Lehane MJ, Wu D, Lehane SM: **Midgut-specific immune molecules are produced by**
1364 **the blood-sucking insect *Stomoxys calcitrans*.** *Proc Natl Acad Sci USA* 1997,
1365 **94:**11502-11507.
1366 125. Mellroth P, Karlsson J, Steiner H: **A scavenger function for a *Drosophila***
1367 **peptidoglycan recognition protein.** *J Biol Chem* 2003, **278:**7059-7064.
1368 126. Bischoff V, Vignal C, Duvic B, Boneca IG, Hoffmann JA, Royet J: **Downregulation of**
1369 **the *Drosophila* immune response by peptidoglycan-recognition proteins SC1 and**
1370 **SC2.** *PLoS Pathog* 2006, **2:**139-147.
1371 127. Gao YF, Tang T, Gu JH, Sun LL, Gao XB, Ma XY, Wang XC, Liu FS, Wang JH:
1372 **Downregulation of the *Musca domestica* peptidoglycan recognition protein SC**
1373 **(PGRP-SC) leads to overexpression of antimicrobial peptides and tardy pupation.**
1374 *Mol Immunol* 2015, **67:**465-474.
1375 128. Ribeiro JMC: **Blood-feeding arthropods - Live syringes or invertebrate**
1376 **pharmacologists.** *Infect Agent Dis* 1995, **4:**143-152.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1377 129. Ribeiro JMC, Arca B: **From sialomes to the sialoverse: An insight into salivary**
1378 **potion of blood-feeding insects.** *Adv Insect Physiol* 2009, **37:**59-118.
1379 130. Ribeiro JMC, Topalis P, Louis C: **AnoXcel: an *Anopheles gambiae* protein database.**
1380 *Insect Mol Biol* 2004, **13:**449-457.
1381 131. Wang XY, Ribeiro JMC, Broce AB, Wilkerson MJ, Kanost MR: **An insight into the**
1382 **transcriptome and proteome of the salivary gland of the stable fly, *Stomoxys***
1383 ***calcitrans*.** *Insect Biochem Mol Biol* 2009, **39:**607-614.
1384 132. Ameri M, Wang X, Wilkerson MJ, Kanost MR, Broce AB: **An immunoglobulin binding**
1385 **protein (Antigen 5) of the stable fly (Diptera : Muscidae) salivary gland stimulates**
1386 **bovine immune responses.** *J Med Entomol* 2008, **45:**94-101.
1387 133. Zhang D, Cupp MS, Cupp EW: **Thrombostasin: purification, molecular cloning and**
1388 **expression of a novel anti-thrombin protein from horn fly saliva.** *Insect Biochem*
1389 *Mol Biol* 2002, **32:**321-330.
1390 134. Ribeiro JMC, Francischetti IMB: **Role of arthropod saliva in blood feeding: Sialome**
1391 **and post-sialome perspectives.** *Annu Rev Entomol* 2003, **48:**73-88.
1392 135. Ribeiro JMC, Mans BJ, Arca B: **An insight into the sialome of blood-feeding**
1393 **Nematocera.** *Insect Biochem Mol Biol* 2010, **40:**767-784.
1394 136. Takac P, Nunn MA, Meszaros J, Pechanova O, Vrbjar N, Vlasakova P, Kozanek M,
1395 Kazimirova M, Hart G, Nuttall PA, Labuda M: **Vasotab, a vasoactive peptide from**
1396 **horse fly *Hybomitra bimaculata* (Diptera, Tabanidae) salivary glands.** *J Exp Biol*
1397 2006, **209:**343-352.
1398 137. Niwa R, Niwa YS: **Enzymes for ecdysteroid biosynthesis: their biological functions**
1399 **in insects and beyond.** *Biosci Biotech Bioch* 2014, **78:**1283-1292.
1400 138. Schuler MA, Berenbaum MR: **Structure and function of cytochrome P450S in insect**
1401 **adaptation to natural and synthetic toxins: Insights gained from molecular**
1402 **modeling.** *J Chem Ecol* 2013, **39:**1232-1245.
1403 139. Barnes KG, Weedall GD, Ndula M, Irving H, Mzihalowa T, Hemingway J, Wondji CS:
1404 **Genomic footprints of selective sweeps from metabolic resistance to pyrethroids**
1405 **in African malaria vectors are driven by scale up of insecticide-based vector**
1406 **control.** *PLoS Genet* 2017, **13**.
1407 140. Le Goff G, Hilliou F: **Resistance evolution in *Drosophila*: the case of CYP6G1.** *Pest*
1408 *Manag Sci* 2017, **73:**493-499.
1409 141. Seifert J, Scott JG: **The CYP6D1v1 allele is associated with pyrethroid resistance in**
1410 **the house fly, *Musca domestica*.** *Pestic Biochem Phys* 2002, **72:**40-44.
1411 142. Feyereisen R: **Arthropod CYPomes illustrate the tempo and mode in P450**
1412 **evolution.** *BBA-Proteins Proteom* 2011, **1814:**19-28.
1413 143. Bloomquist JR: **Chloride channels as tools for developing selective insecticides.**
1414 *Arch Insect Biochem Physiol* 2003, **54:**145-156.
1415 144. Ffrench-Constant RH, Williamson MS, Davies TG, Bass C: **Ion channels as insecticide**
1416 **targets.** *J Neurogenet* 2016, **30:**163-177.
1417 145. Pitzer JB, Kaufman PE, Tenbroeck SH: **Assessing permethrin resistance in the**
1418 **stable fly (Diptera: Muscidae) in Florida by using laboratory selections and field**
1419 **evaluations.** *J Econ Entomol* 2010, **103:**2258-2263.
1420 146. Salem A, Bouhsira E, Liénard E, Melou AB, Jacquiet P, Franc M: **Susceptibility of two**
1421 **European strains of *Stomoxys calcitrans* (L.) to cypermethrin, deltamethrin,**
1422 **fenvalerate, λ-cyhalothrin, permethrin and phoxim.** *Int J Appl Res Vet Med* 2012,
1423 **10:**249-257.
1424 147. Tainchum K, Shukri S, Duvallet G, Etienne L, Jacquiet P: **Phenotypic susceptibility to**
1425 **pyrethroids and organophosphate of wild *Stomoxys calcitrans* (Diptera:**
1426 **Muscidae) populations in southwestern France.** *Parasitol Res* 2018, **117:**4027-4032.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1427 148. Olafson PU, Pitzer JB, Kaufman PE: **Identification of a mutation associated with**
1428 **permethrin resistance in the para-type sodium channel of the stable fly (Diptera:**
1429 **Muscidae).** *J Econ Entomol* 2011, **104:**250-257.
1430 149. Olafson PU, Kaufman PE, Duvallet G, Solorzano JA, Taylor DB, Fryxell RT: **Frequency**
1431 **of *kdr* and *kdr*-his alleles in stable fly (Diptera: Muscidae) populations from the**
1432 **United States, Costa Rica, France, and Thailand.** *J Med Entomol* 2019, ***In Revision***.
1433 150. Tseng JM, Jones CJ, Hogsette JA: **Nectar feeding and the stable fly, *Stomoxys***
1434 ***calcitrans* (Diptera: Muscidae).** *J Fla Anti-Mosq Assoc* 1983, **54:**40-41.
1435 151. Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S: **Vitamin B6 generated by**
1436 **obligate symbionts is critical for maintaining proline homeostasis and fecundity in**
1437 **tsetse flies.** *Appl Environ Microbiol* 2014, **80:**5844-5853.
1438 152. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S: **Genome**
1439 **sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia***
1440 ***glossinidia*.** *Nat Genet* 2002, **32:**402-407.
1441 153. Janda JM, Abbott SL: **The genus Aeromonas: taxonomy, pathogenicity, and**
1442 **infection.** *Clin Microbiol Rev* 2010, **23:**35-73.
1443 154. Wang Y, Gilbreath TM, 3rd, Kukutla P, Yan G, Xu J: **Dynamic gut microbiome across**
1444 **life history of the malaria mosquito Anopheles gambiae in Kenya.** *PLoS One* 2011,
1445 **6:**e24767.
1446 155. Coon KL, Vogel KJ, Brown MR, Strand MR: **Mosquitoes rely on their gut microbiota**
1447 **for development.** *Mol Ecol* 2014, **23:**2727-2739.
1448 156. Guégan M, Zouache K, Démichel C, Minard G, Potier P, Mavingui P, Moro CV: **The**
1449 **mosquito holobiont: fresh insight into mosquito-microbiota interactions.**
1450 *Microbiome* 2018, **6:**49.
1451 157. Tomberlin JK, Crippen TL, Tarone AM, Chaudhury MFB, Singh B, Cammack JA, Meisel
1452 RP: **A review of bacterial interactions with blow flies (Diptera: Calliphoridae) of**
1453 **medical, veterinary, and forensic importance.** *Ann Entomol Soc Amer* 2017, **110:**19-
1454 36.
1455 158. Weatherbee CR, Pechal JL, Benbow ME: **The dynamic maggot mass microbiome.**
1456 *Ann Entomol Soc Amer* 2017, **110:**45-53.
1457 159. Castro BG, Souza MMS, Regua-Mangia AH, Bittencourt AJ: **Enterobacterial**
1458 **microbiota on *Stomoxys calcitrans* external surface.** *Transbound Emerg Dis* 2010,
1459 **57:**22-24.
1460 160. de Castro BG, de Souza MMS, Regua-Mangia AH, Bittencourt AJ: **Occurrence of**
1461 **Shiga-toxigenic *Escherichia coli* in *Stomoxys calcitrans* (Diptera: Muscidae).** *Rev*
1462 *Bras Parasitol V* 2013, **22:**318-321.
1463 161. Werren JH, Baldo L, Clark ME: **Wolbachia: master manipulators of invertebrate**
1464 **biology.** *Nat Rev Microbiol* 2008, **6:**741-751.
1465 162. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH: **How many**
1466 **species are infected with Wolbachia? - a statistical analysis of current data.** *Fems*
1467 *Microbiol Lett* 2008, **281:**215-220.
1468 163. Zug R, Hammerstein P: **Still a host of hosts for *Wolbachia*: Analysis of recent data**
1469 **suggests that 40% of terrestrial arthropod species are infected.** *PLoS One* 2012, **7**.
1470 164. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi
1471 HS, Lambert SA, Mann I, Cook K: **Determination and inference of eukaryotic**
1472 **transcription factor sequence specificity.** *Cell* 2014, **158:**1431-1443.
1473 165. Kaufman TC, Seeger MA, Olsen G: **Molecular and genetic organization of the**
1474 ***antennapedia* gene complex of *Drosophila melanogaster*.** *Adv Genet* 1990, **27:**309-
1475 362.

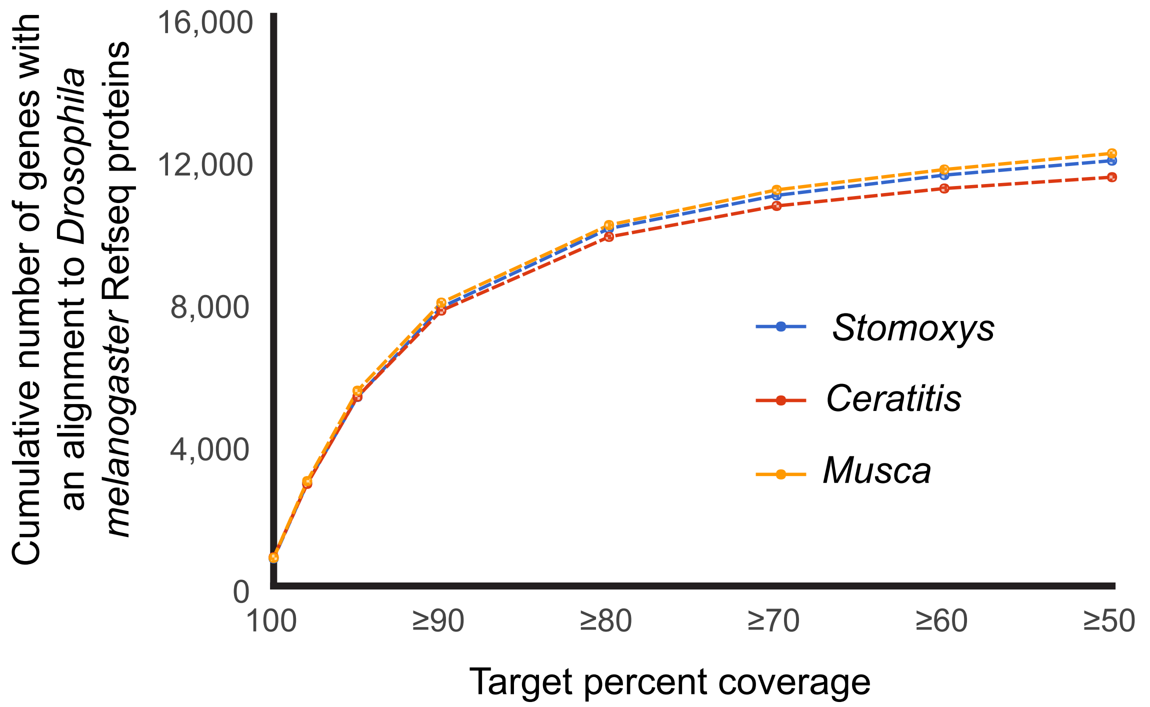
bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

1476 166. Simeone A, Dapice MR, Nigro V, Casanova J, Graziani F, Acampora D, Avantaggiato V:
1477 ***Orthopedia*, a novel homeobox-containing gene expressed in the developing CNS**
1478 **of both mouse and *Drosophila*.** *Neuron* 1994, **13:**83-101.
1479 167. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall
1480 G, Shea TP, Sykes S, et al: **High-quality draft assemblies of mammalian genomes**
1481 **from massively parallel sequence data.** *Proc Natl Acad Sci U S A* 2011, **108:**1513-
1482 1518.
1483 168. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: **Scaffolding pre-assembled**
1484 **contigs using SSPACE.** *Bioinformatics* 2011, **27:**578-579.
1485 169. Rosendale AJ, Romick-Rosendale LE, Watanabe M, Dunlevy ME, Benoit JB:
1486 **Mechanistic underpinnings of dehydration stress in ticks revealed through RNA-**
1487 **seq and metabolomics.** *J Exp Biol* 2016, **219** 1808-1819.
1488 170. Benoit JB, Attardo GM, Michalkova V, Krause TB, Bohova J, Zhang Q, Baumann AA,
1489 Mireji PO, Takac P, Denlinger DL, et al: **A novel highly divergent protein family**
1490 **identified from a viviparous insect by RNA-seq analysis: a potential target for**
1491 **tsetse fly-specific abortifacients.** *PLoS Genet* 2014, **10:**e1003874.
1492 171. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J: **g:Profiler-a web**
1493 **server for functional interpretation of gene lists (2016 update).** *Nucleic Acids Res*
1494 2016, **44:**W83-W89.
1495 172. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H,
1496 Glass CK: **Simple combinations of lineage-determining transcription factors prime**
1497 **cis-regulatory elements required for macrophage and B cell identities.** *Mol Cell*
1498 2010, **38:**576-589.
1499 173. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA,
1500 Oakley BB, Parks DH, Robinson CJ, et al: **Introducing mothur: Open-source,**
1501 **platform-independent, community-supported software for describing and**
1502 **comparing microbial communities.** *Appl Environ Microbiol* 2009, **75:**7537-7541.
1503 174. Wheeler D, Redding AJ, Werren JH: **Characterization of an ancient Lepidopteran**
1504 **lateral gene transfer.** *PLoS One* 2013, **8.**
1505 175. Poynton HC, Hasenbein S, Benoit JB, Sepulveda MS, Poelchau MF, Hughes DST,
1506 Murali SC, Chen S, Glastad KM, Goodisman MAD, et al: **The toxicogenome of**
1507 ***Hyalella azteca*: A model for sediment ecotoxicology and evolutionary toxicology.**
1508 *Environ Sci Technol* 2018, **52:**6009-6022.
1509 176. Panfilio KA, Vargas Jentzsch IM, Benoit JB, Erezyilmaz D, Suzuki Y, Colella S,
1510 Robertson HM, Poelchau MF, Waterhouse RM, Ioannidis P, et al: **Molecular**
1511 **evolutionary trends and feeding ecology diversification in the Hemiptera,**
1512 **anchored by the milkweed bug genome.** *Genome Biol* 2019, **20:**64.
1513 177. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ: **W-IQ-TREE: a fast online**
1514 **phylogenetic tool for maximum likelihood analysis.** *Nucleic Acids Res* 2016,
1515 **44:**W232-235.
1516 178. Salcedo E, Huber A, Henrich S, Chadwell LV, Chou WH, Paulsen R, Britt SG: **Blue- and**
1517 **green-absorbing visual pigments of *Drosophila*: Ectopic expression and**
1518 **physiological characterization of the R8 photoreceptor cell-specific Rh5 and Rh6**
1519 **rhodopsins.** *J Neurosci* 1999, **19:**10716-10726.
1520 179. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high**
1521 **throughput.** *Nucleic Acids Res* 2004, **32:**1792-1797.
1522 180. Castresana J: **Selection of conserved blocks from multiple alignments for their use**
1523 **in phylogenetic analysis.** *Mol Biol Evol* 2000, **17:**540-552.
1524 181. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: Molecular**
1525 **Evolutionary Genetics Analysis Version 6.0.** *Mol Biol Evol* 2013, **30:**2725-2729.

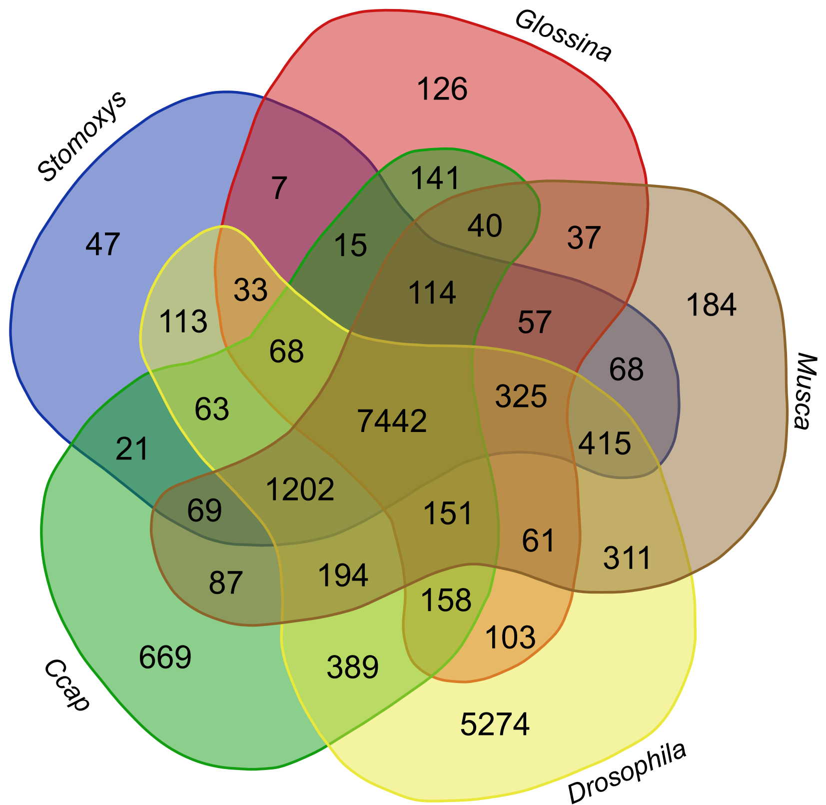
bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

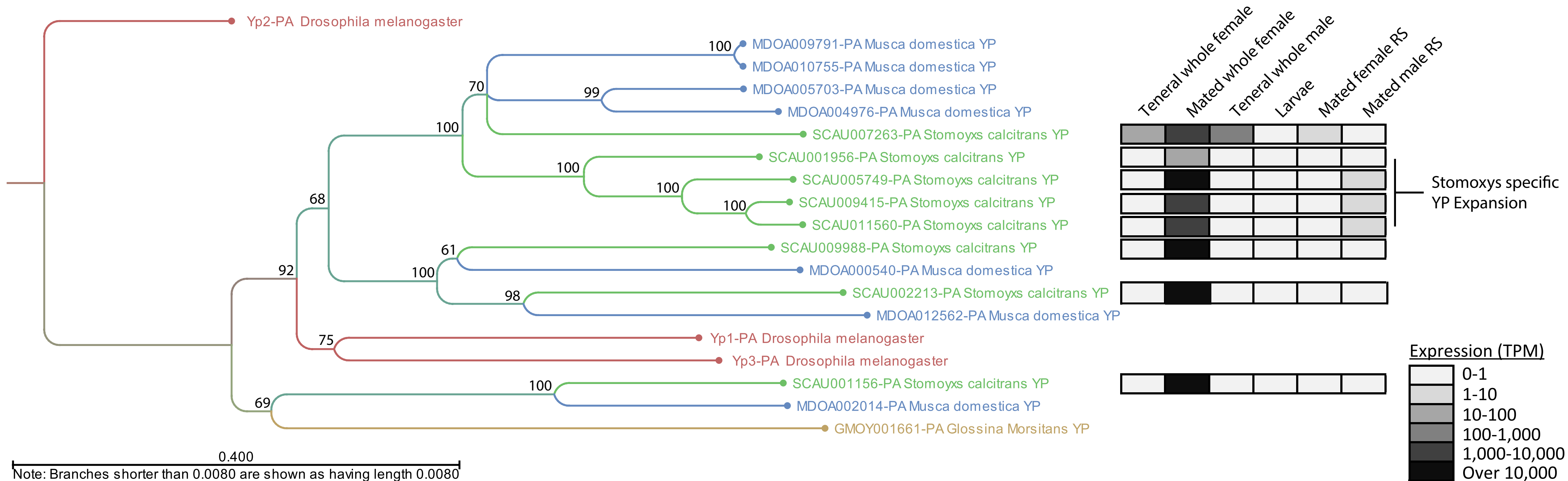
1526 182. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated**
1527 **alignment trimming in large-scale phylogenetic analyses.** *Bioinformatics* 2009,
1528 **25:**1972-1973.
1529

A.

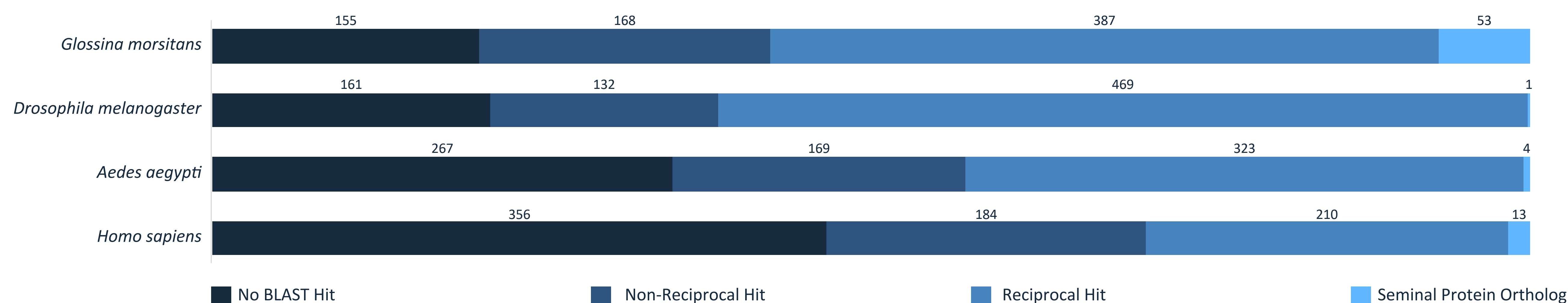


B.

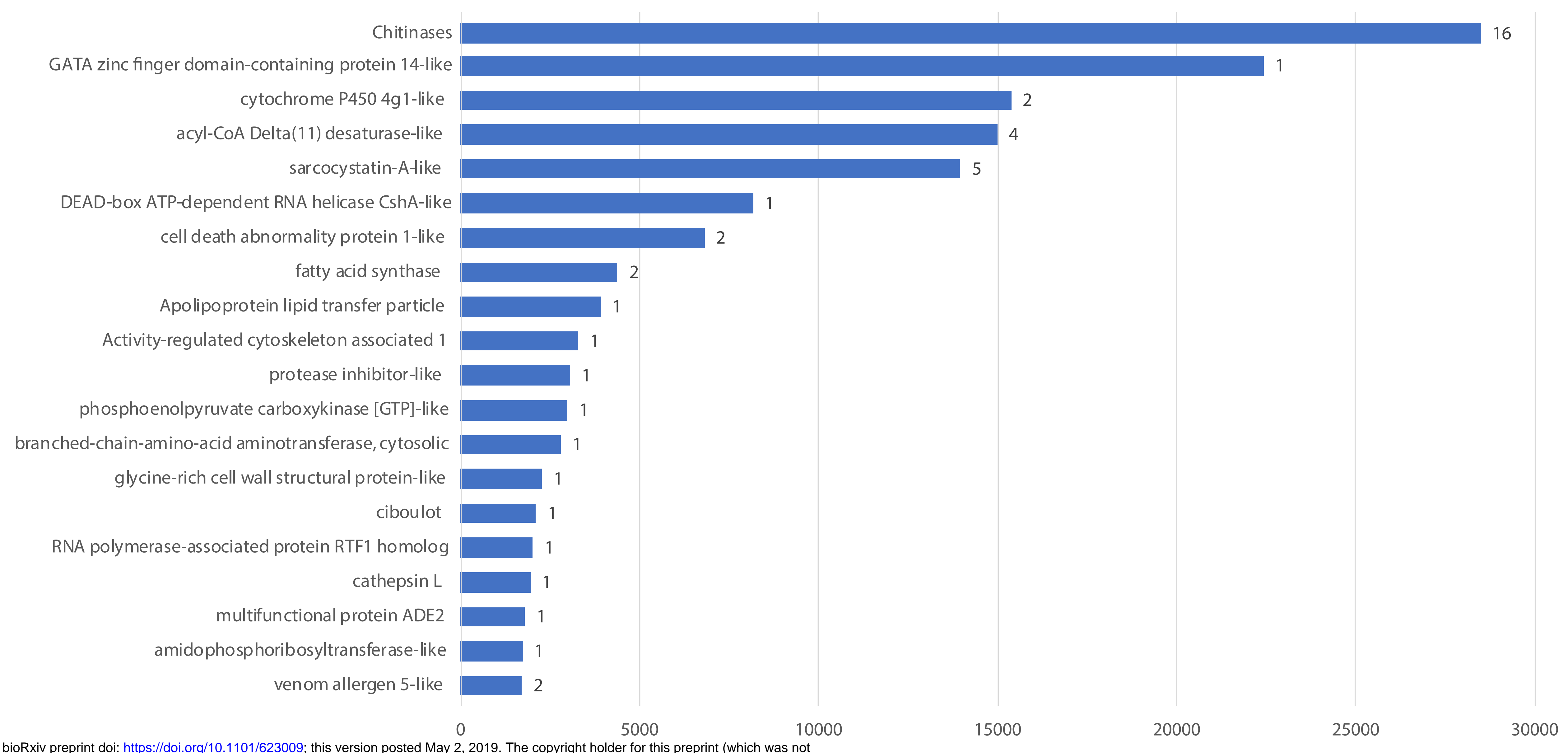



Yp2-PA Drosophila melanogaster
MDOA009791-PA Musca domestica YP
MDOA010755-PA Musca domestica YP
MDOA005703-PA Musca domestica YP
MDOA004976-PA Musca domestica YP
SCAU007263-PA Stomoyxs calcitrans YP
SCAU001956-PA Stomoyxs calcitrans YP
SCAU005749-PA Stomoyxs calcitrans YP
SCAU009415-PA Stomoyxs calcitrans YP
SCAU011560-PA Stomoyxs calcitrans YP
SCAU009988-PA Stomoyxs calcitrans YP
MDOA000540-PA Musca domestica YP
SCAU002213-PA Stomoyxs calcitrans YP
MDOA012562-PA Musca domestica YP
Yp1-PA Drosophila melanogaster
Yp3-PA Drosophila melanogaster
SCAU001156-PA Stomoyxs calcitrans YP
MDOA002014-PA Musca domestica YP
GMOY001661-PA Glossina Morsitans YP
100
99
70
100
100
100
100
68
61
100
92
98
75
100
69
Teneral whole female
Mated whole female
Teneral whole male
Larvae
Mated female RS
Mated male RS
Stomoxys specific
YP Expansion
Expression (TPM)
0-1
1-10
10-100
100-1,000
1,000-10,000
Over 10,000
0.400
Note: Branches shorter than 0.0080 are shown as having length 0.0080

A



B



bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

C

Male Reproductive Tract Expression Values (Log2)

Male/Female Fold Change (Log 2)

Signal Peptide Status: No Signal Peptide; Signal Peptide

Seminal Protein Orthology: Non−Seminal Protein; Seminal Protein

Log2 Fold Change: 4, 6, 8, 10, 12

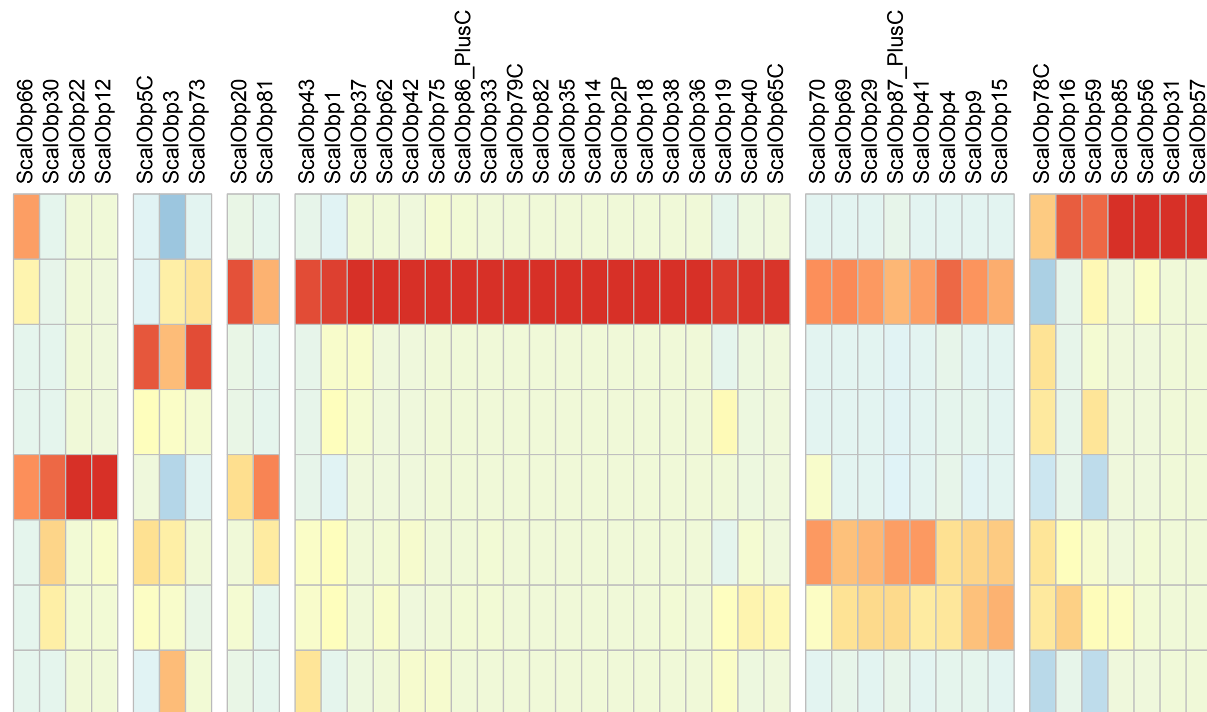
Log2 Male Expression: 6, 8, 10, 12, 14

A

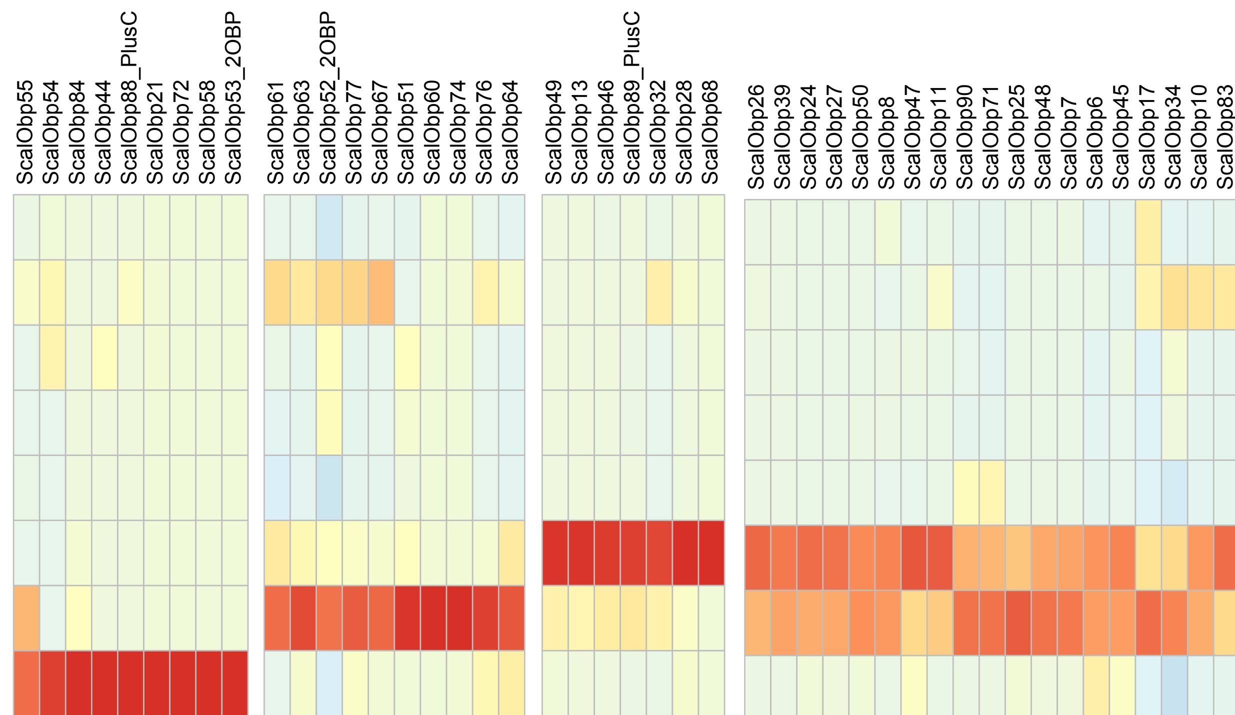
0.5

B

Larvae
Mated Male Reproductive System
Teneral Whole Male
Teneral Whole Female
Mated Female Reproductive System
Mated Male Head
Mated Female Head
Mated Whole Female

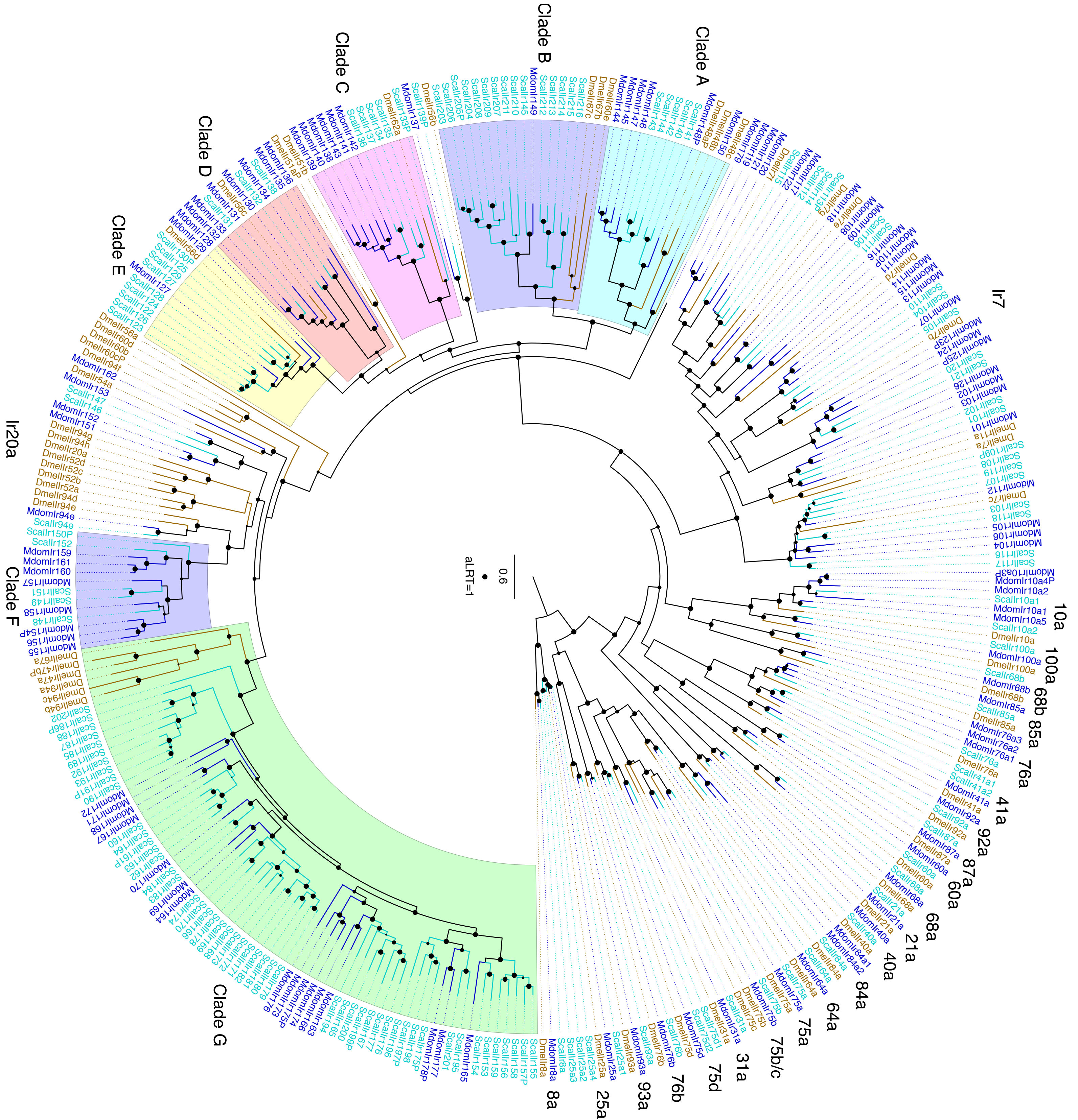


Larvae
Mated Male Reproductive System
Teneral Whole Male
Teneral Whole Female
Mated Female Reproductive System
Mated Male Head
Mated Female Head
Mated Whole Female



bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



a.

Opsin subfamily: Rh7 UV B LW

Visual organ: BR COMPOUND EYE OC

Rh7 Rh3 Rh4 Rh5 Rh6 Rh1 Rh2

Drosophila melanogaster 420 325 375 437 508 480 420

bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Drosophila virilis

Musca domestica 1.1.2 1.1.1.1 1.1.1.2

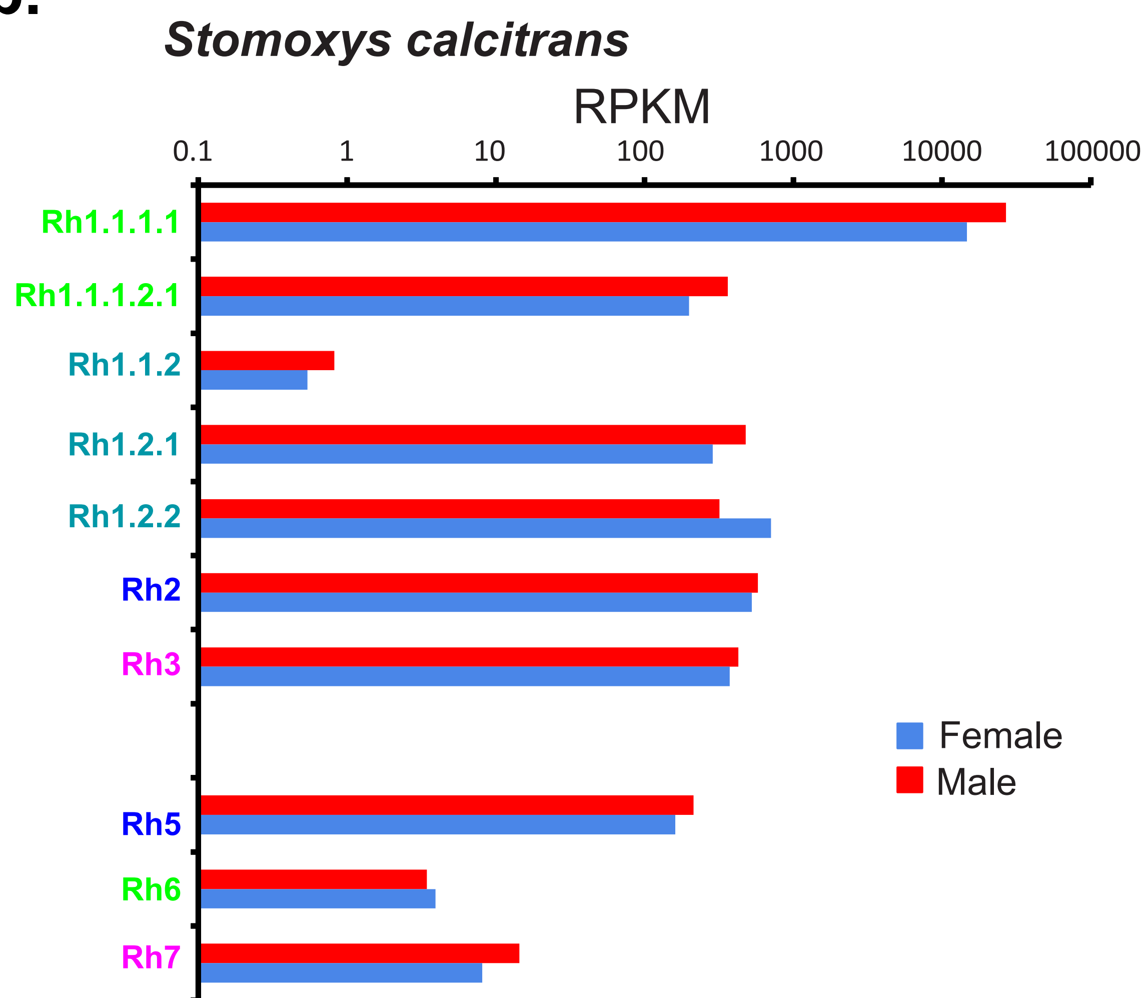
Stomoxys calcitrans 1.1.2 1.1.1.1 1.1.1.2.1 1.1.1.2.2 1.2.1 1.2.2

Glossina morsitans

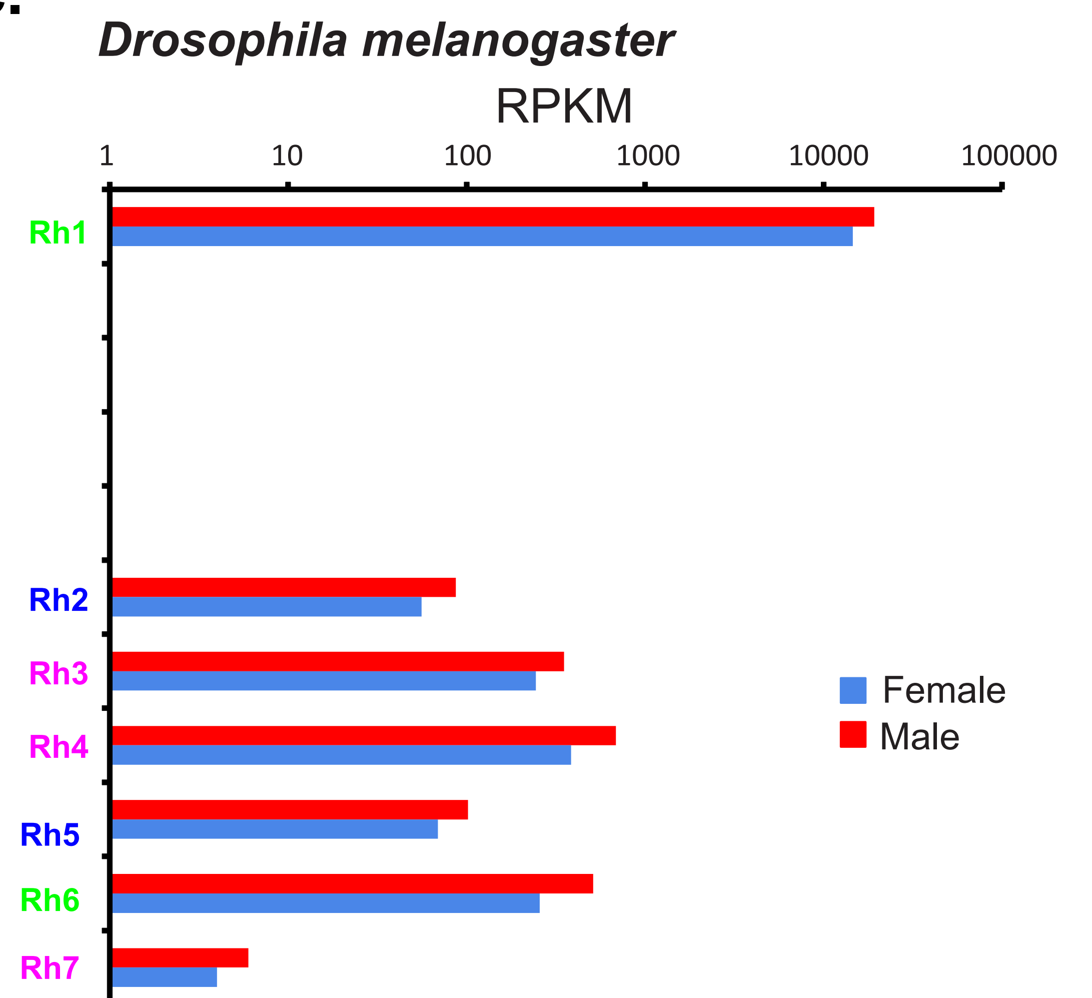
Ceratitis capitata

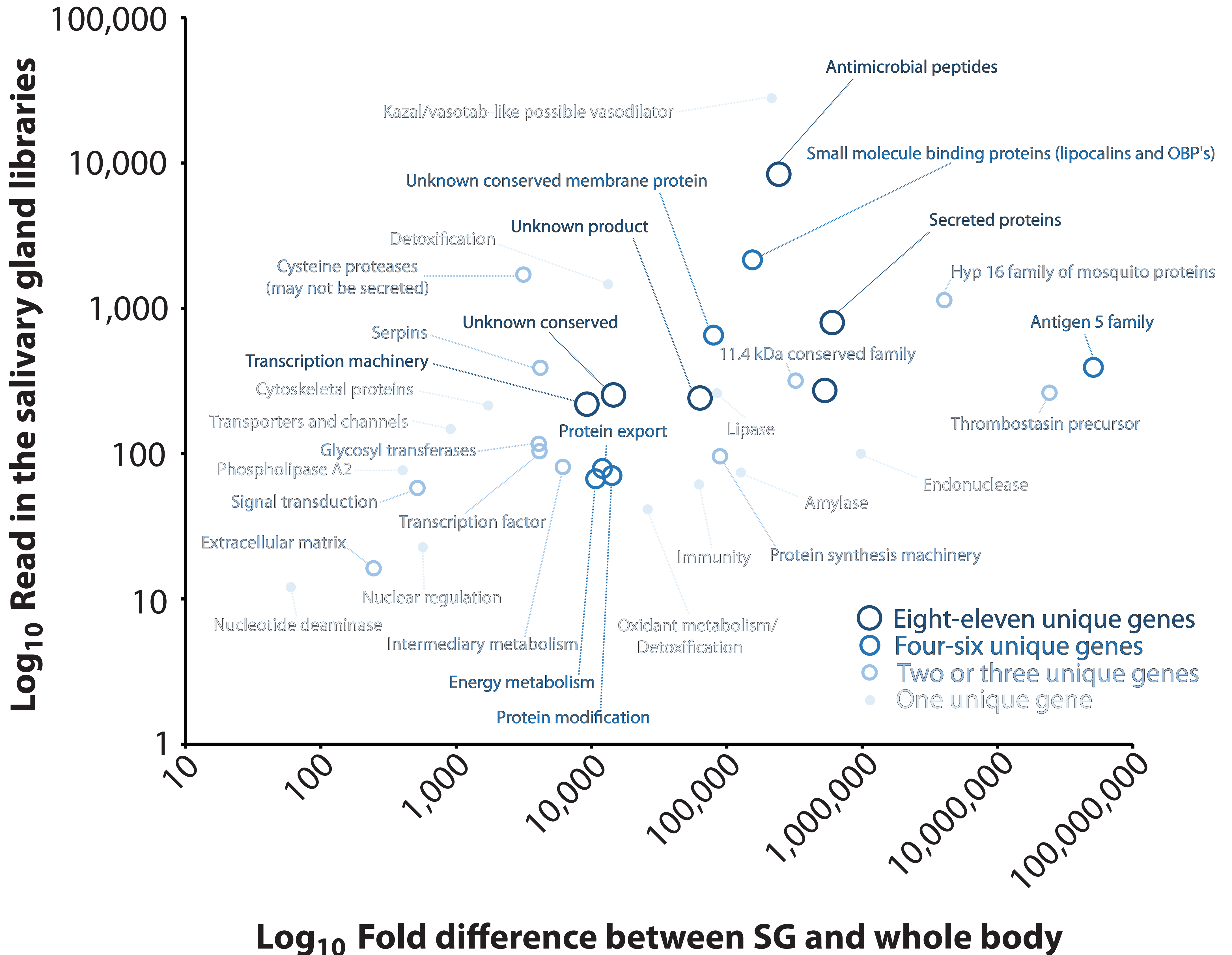
100 50 0 Mya

b.



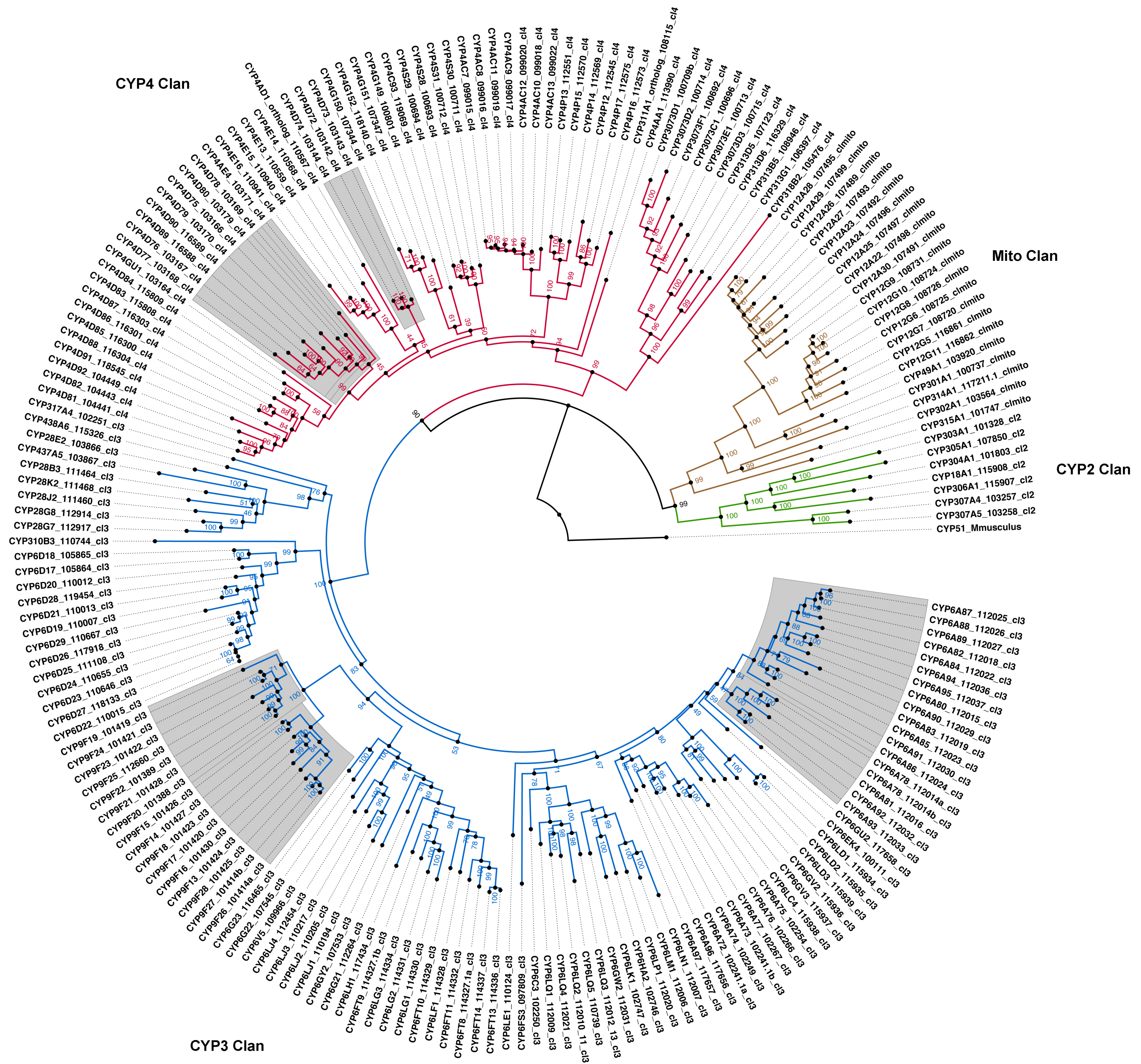
c.



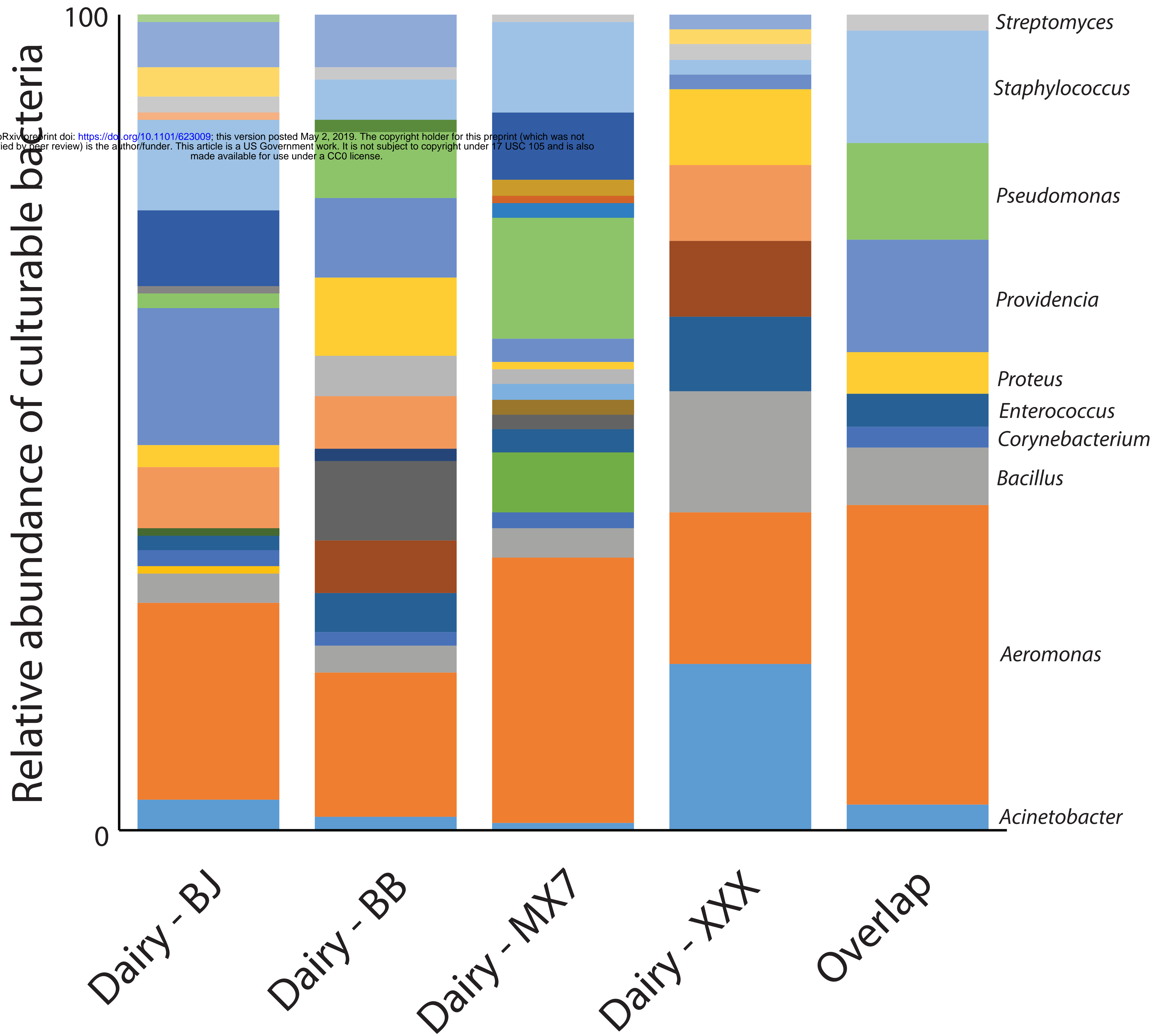


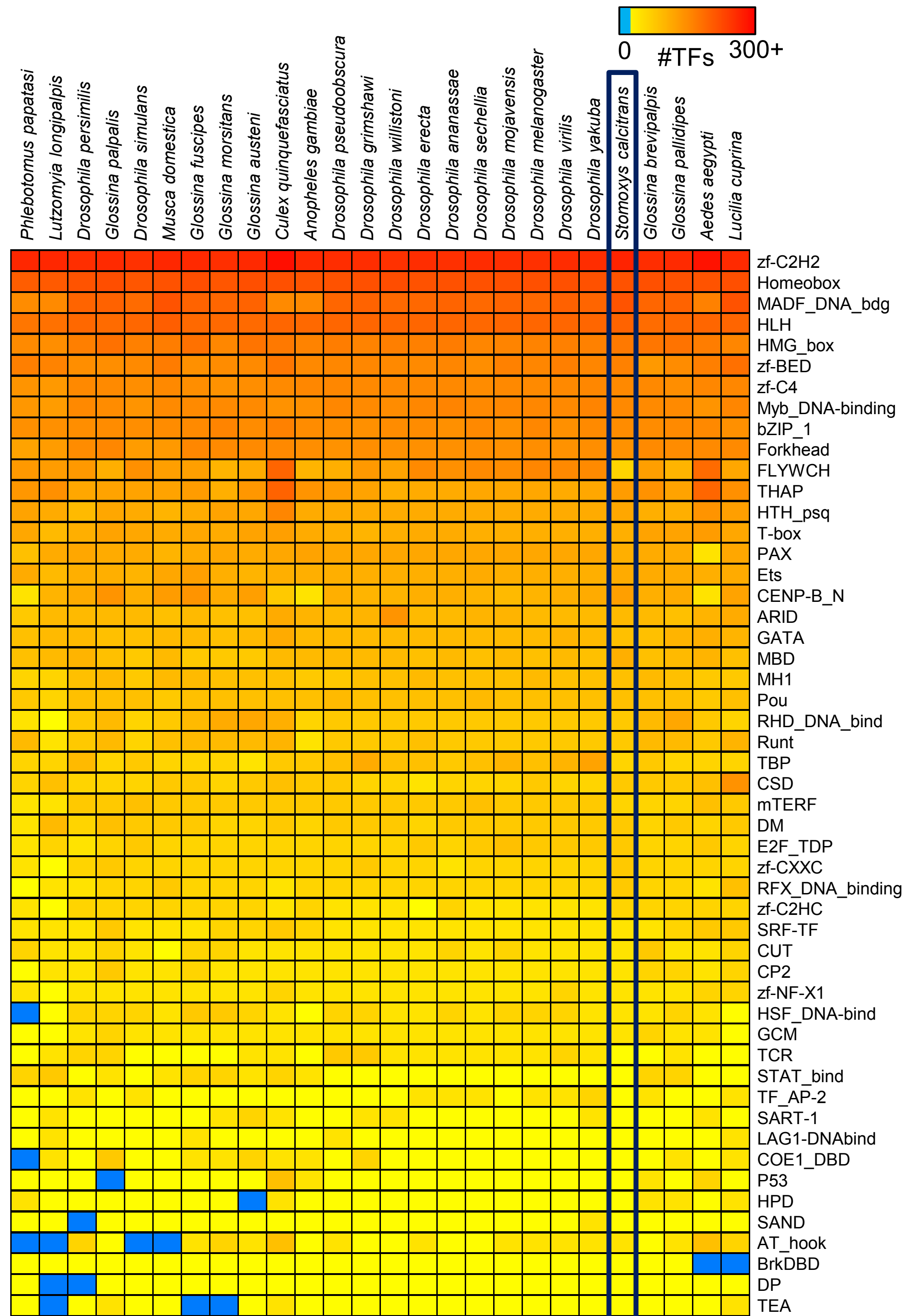
Log10 Read in the salivary gland libraries
100,000
10,000
1,000
100
10
1
Antimicrobial peptides
Kazal/vasotab-like possible vasodilator
Small molecule binding proteins (lipocalins and OBP's)
Unknown conserved membrane protein
Secreted proteins
Unknown product
Detoxification
Cysteine proteases
(may not be secreted)
Hyp 16 family of mosquito proteins
Unknown conserved
Antigen 5 family
Serpins
11.4 kDa conserved family
Transcription machinery
Cytoskeletal proteins
Thrombostasin precursor
Transporters and channels
Lipase
Protein export
Glycosyl transferases
Phospholipase A2
Endonuclease
Signal transduction
Amylase
Transcription factor
Extracellular matrix
Immunity
Protein synthesis machinery
Nuclear regulation
Nucleotide deaminase
Oxidant metabolism/
Detoxification
Intermediary metabolism
Energy metabolism
Protein modification
Eight-eleven unique genes
Four-six unique genes
Two or three unique genes
One unique gene
10
100
1,000
10,000
100,000
1,000,000
10,000,000
100,000,000
Log10 Fold difference between SG and whole body

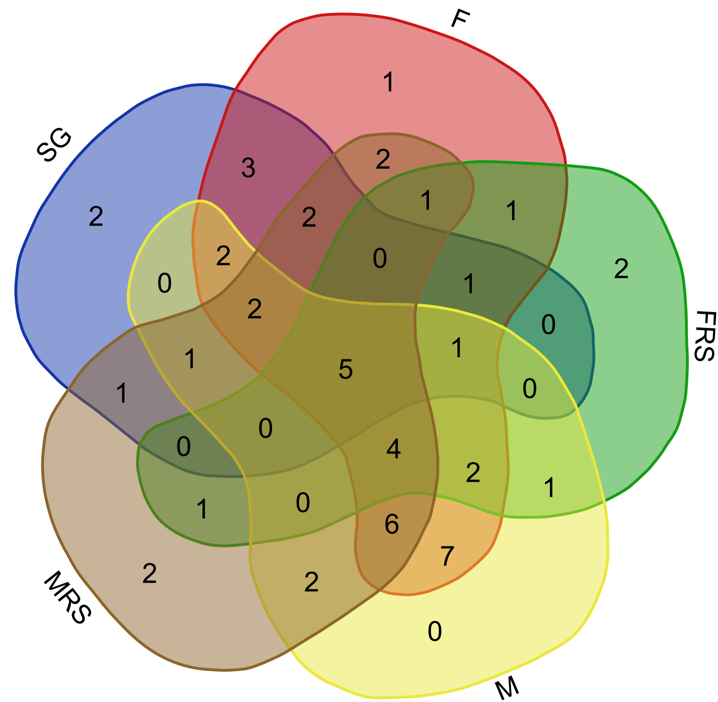
bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

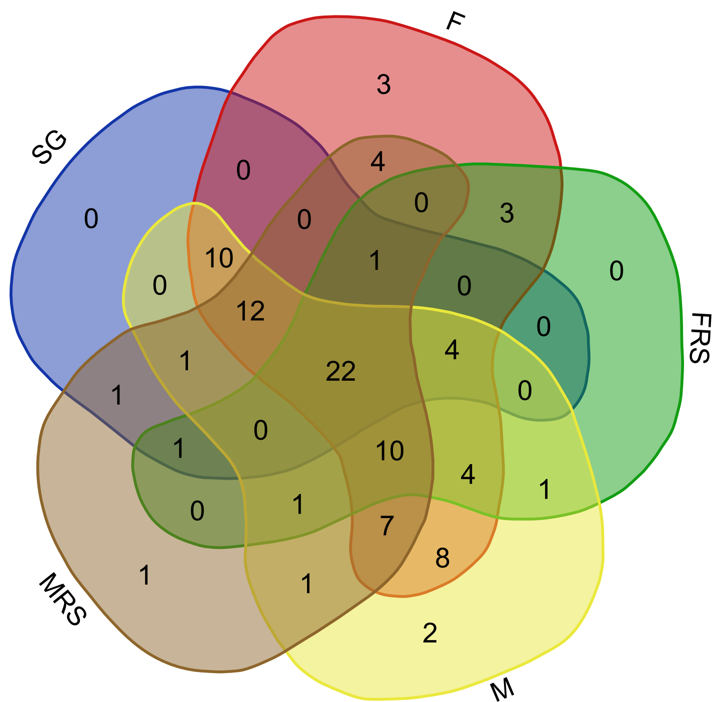


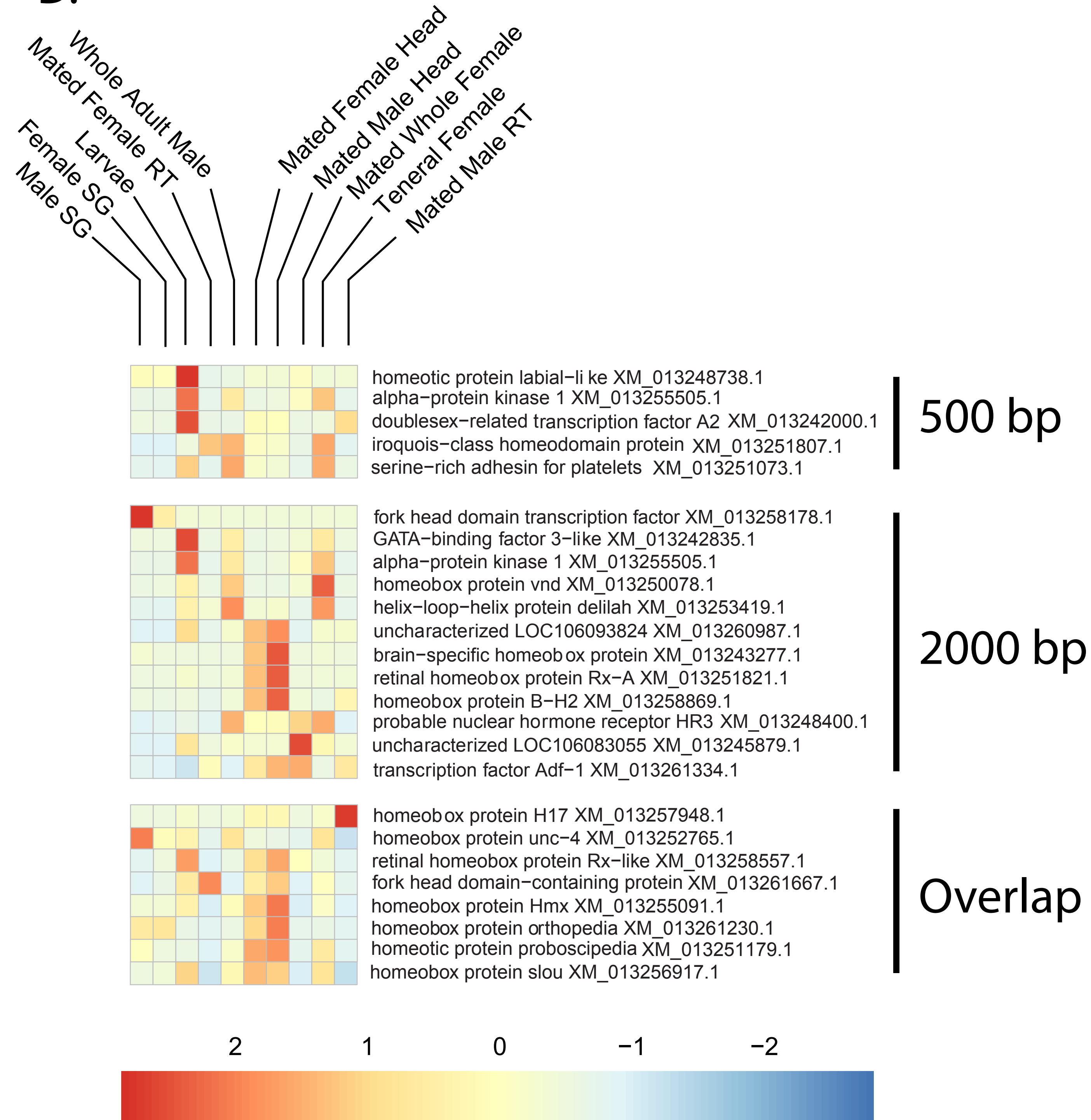
bioRxiv preprint doi: https://doi.org/10.1101/623009; this version posted May 2, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.




A.
0
#TFs
300+
Phlebotomus papatasi
Lutzomyia longipalpis
Drosophila persimilis
Glossina palpalis
Drosophila simulans
Musca domestica
Glossina fuscipes
Glossina morsitans
Glossina austeni
Culex quinquefasciatus
Anopheles gambiae
Drosophila pseudoobscura
Drosophila grimshawi
Drosophila willistoni
Drosophila erecta
Drosophila ananassae
Drosophila sechellia
Drosophila mojavensis
Drosophila melanogaster
Drosophila virilis
Drosophila yakuba
Stomoxys calcitrans
Glossina brevipalpis
Glossina pallidipes
Aedes aegypti
Lucilia cuprina
zf-C2H2
Homeobox
MADF_DNA_bdg
HLH
HMG_box
zf-BED
zf-C4
Myb_DNA-binding
bZIP_1
Forkhead
FLYWCH
THAP
HTH_psq
T-box
PAX
Ets
CENP-B_N
ARID
GATA
MBD
MH1
Pou
RHD_DNA_bind
Runt
TBP
CSD
mTERF
DM
E2F_TDP
zf-CXXC
RFX_DNA_binding
zf-C2HC
SRF-TF
CUT
CP2
zf-NF-X1
HSF_DNA-bind
GCM
TCR
STAT_bind
TF_AP-2
SART-1
LAG1-DNAbind
COE1_DBD
P53
HPD
SAND
AT_hook
BrkDBD
DP
TEA


B. 500 bp
SG
F
FRS
M
MRS


C. 2000 bp
SG
F
FRS
M
MRS


D.
Male SG
Female SG
Larvae
Mated Female RT
Whole Adult Male
Mated Female Head
Mated Male Head
Mated Whole Female
Teneral Female
Mated Male RT
homeotic protein labial−li ke XM_013248738.1
alpha−protein kinase 1 XM_013255505.1
doublesex−related transcription factor A2 XM_013242000.1
iroquois−class homeodomain protein XM_013251807.1
serine−rich adhesin for platelets XM_013251073.1
500 bp
fork head domain transcription factor XM_013258178.1
GATA−binding factor 3−like XM_013242835.1
alpha−protein kinase 1 XM_013255505.1
homeobox protein vnd XM_013250078.1
helix−loop−helix protein delilah XM_013253419.1
uncharacterized LOC106093824 XM_013260987.1
brain−specific homeob ox protein XM_013243277.1
retinal homeob ox protein Rx−A XM_013251821.1
homeobox protein B−H2 XM_013258869.1
probable nuclear hormone receptor HR3 XM_013248400.1
uncharacterized LOC106083055 XM_013245879.1
transcription factor Adf−1 XM_013261334.1
2000 bp
homeob ox protein H17 XM_013257948.1
homeobox protein unc−4 XM_013252765.1
retinal homeobox protein Rx−like XM_013258557.1
fork head domain−containing protein XM_013261667.1
homeobox protein Hmx XM_013255091.1
homeobox protein orthopedia XM_013261230.1
homeotic protein proboscipedia XM_013251179.1
homeobox protein slou XM_013256917.1
Overlap
2
1
0
−1
−2